Andreas Martin Bischoff

Eo Ipso – Automated Internal Reconstruction

TLP-Dissertationen/Sprachwissenschaften

EO IPSO –
AUTOMATED INTERNAL RECONSTRUCTION

Andreas Martin Bischoff

EO IPSO – AUTOMATED INTERNAL RECONSTRUCTION

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

TÜBINGEN
LIBRARY PUBLISHING

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie, detaillierte bibliografische Daten sind im Internet über http://dnb.dnb.de abrufbar.

Die Online-Version dieser Publikation ist auf dem Repositorium der Universität Tübingen frei verfügbar (Open Access).
http://hdl.handle.net/10900/133321
http://nbn-resolving.de/urn:nbn:de:bsz:21-dspace-1333214
http://dx.doi.org/10.15496/publikation-74674

Tübingen Library Publishing 2023
Universitätsbibliothek Tübingen
Wilhelmstraße 32
72074 Tübingen
druckdienste@ub.uni-tuebingen.de
https://tlp.uni-tuebingen.de

ISBN (Softcover): 978-3-946552-73-4
ISBN (PDF): 978-3-946552-74-1

Umschlaggestaltung: Sandra Binder, Universitätsbibliothek Tübingen
Satz: Andreas Martin Bischoff
Druck und Bindung: Druckhaus Sportflieger in der medialis Offsetdruck GmbH
Printed in Germany

Abstract

The doctoral thesis deals with the automation of internal reconstruction of ancient language stages. In the first part of the work, it begins with a discussion of the theoretical aspects of internal reconstruction, as well as its problems and basic requirements. Different methods that have been proposed in the literature are presented and their suitability for an adequate computational implementation is discussed. The procedure and the evaluation of the implemented methods are the subject of the second part of the thesis. The emphasis is set on the phonological reconstruction of sound change. In total, six different methods of internal reconstruction are implemented and evaluated with German and Proto-Indo-European data. The paradigmatic, derivational, and semantic methods can be counted among the morphophonemic methods of internal reconstruction. They assume that two alternating wordforms go back to a historical *pre-form* and that the morphophonemic alternation was caused by a sound change. These wordforms are either inflectional forms of the same paradigm or derivations, or they belong to the same semantic field. Three other implemented methods are based on the distributions of phons and phonemes in a language. The phonotactic method considers the frequency of phonotagms in a corpus or word list to be the result of preceding sound changes, while the distinctive method tries to infer sound changes by means of minimal pairs. Finally, the gap approach derives historical mergers from gaps within a phoneme system by assuming symmetries of the phoneme system.

The evaluation found that the derivational and paradigmatic methods achieve the best result with a precision of 0.5238 for the derivational method and precision of 0.4 for the paradigmatic method among the first 25 automatically identified sound correspondences of German. Overall, among the first 100 sound correspondences, 18 (derivational method) and 17 sound pairs (paradigmatic method) can be assigned to historical sound changes. The highest F-scores (0.1217 and 0.1118) are reached after 83 and 39 sound correspondences, respectively. Significantly lower is the highest F-scores of the distributional methods with values of 0.0611 (phonotactic method), 0.0611 (distinctive method), and 0.0581 (gap approach). They each detect only seven sound changes from the used German gold standard. Evidently higher F-score values are obtained in the detection of phonological rules. The paradigmatic

method obtains an F-score of 0.3226, the derivational method 0.3607, the phonotactic method 0.1702, and the distinctive method 0.1538. Only the gap approach cannot reconstruct any phonological rule from the gold standard. This result strongly suggests that the current methods of internal reconstruction are designed for the reconstruction of phonological rules rather than historical sound changes. With respect to the age of the detected sound changes, a tendency towards younger language stages can be observed, but rather, the decisive factor is the dominance of alternations or phonotagms in the corpus. An example of morphological internal reconstruction is given by the semantic method. From the comparison of word pairs which represent the same semantic relation, significant phonemes and phonotagms can be determined that may indicate an unproductive derivational affix of the *pre-language*.

Zusammenfassung

Die Dissertation befasst sich mit der Automatisierung der Methoden der internen Rekonstruktion von antiken Sprachstufen. Sie beginnt im ersten Teil der Arbeit mit einer Erörterung des theoretischen Aspekts der internen Rekonstruktion, sowie dessen Probleme und Grundvoraussetzungen. Unterschiedliche in der Fachliteratur vorgebrachte Methoden werden zunächst erörtert und ihre Eignung für eine adäquate maschinelle Implementierung daraufhin reflektiert. Das Vorgehen und die Evaluation der umgesetzten Methoden sind Inhalt des zweiten Teils der Arbeit. Dabei wird der Schwerpunkt auf die phonologische Rekonstruktion von Lautwandel gesetzt. Insgesamt werden sechs verschiedene interne Rekonstruktionsmethoden implementiert, deren Evaluation mit Daten aus dem Deutschen und dem Urindogermanischen erfolgt. Die paradigmatische, derivationale und semantische Methoden können zu den morphophonemischen Methoden der internen Rekonstruktion gezählt werden. Sie gehen davon aus, dass zwei alternierende Wortformen auf eine historische *Vorform* zurückgehen und die morphophonemische Alternation durch einen Lautwandel verursacht wurde. Diese Wortformen stellen hierbei entweder Formen eines Flexionsparadigmas oder Derivationen dar oder gehören zum selben semantischen Wortfeld. Die drei weiteren umgesetzten Methoden beruhen auf den Distributionen von Phonen und Phonemen in einer Sprache. Die phonotaktische Methode betrachtet die Häufigkeit von Phonotagmen in einem Korpus oder Wortliste als das Ergebnis vorausgegangener Lautwandel, während die distinktive Methode Lautwandel auf Basis von Minimalpaaren zu erschließen versucht. Die *gap approach* leitet schließlich aus Lücken innerhalb eines Phonemsystems ehemals zusammengefallene Laute ab, in dem sie Symmetrien des Phonemsystems konstruiert.

In der Evaluierung erreichen die derivationale und paradigmatische Methoden die besten Ergebnisse und erzielen mit deutschen Testdaten innerhalb der ersten 25 automatisch identifizierten Lautkorrespondenzen eine Präzision von 0,5238 für die derivationale Methode und eine Präzision von 0,4 für die paradigmatische Methode. Insgesamt können von den ersten 100 Lautpaaren 18 (derivationale Methode) bzw. 17 (paradigmatische Methode) einem historischen Lautwandel zugeordnet werden. Der höchste F-score wird mit 0,1217 bzw. 0,1118 nach 83 bzw. 39 Lautkorrespon-

denzen erreicht. Deutlich darunter liegen die besten F-Score-Werte der distributionellen Methoden mit 0,0611 (phonotaktische Methode), 0,0611 (distinktive Methode) und 0.0581 (*gap approach*). Sie erkennen jeweils nur sieben korrekte Lautwandel aus dem herangezogenen deutschen Goldstandard. Evident höhere F-Score-Werte werden bei der Erkennung von phonologischen Regeln erlangt. Der höchste F-Score liegt hier mit der paradigmatischen Methode bei 0,3226, mit der derivationalen Methode bei 0,3607, mit der phonotaktischen Methode bei 0,1702 und mit der distinktiven Methode bei 0,1538. Nur die *gap approach* kann keine phonologische Regel aus dem Goldstandard rekonstruieren. Dieses Ergebnis spricht deutlich dafür, dass die gängigen internen Rekonstruktionsmethoden eher für die Rekonstruktion von phonologischen Regeln als historische Lautwandel konzipiert sind. In Bezug auf das Alter der erkannten Lautwandel zeichnet sich mitunter eine Tendenz zu jüngeren Sprachstufen ab, doch ist hier vielmehr die Dominanz der Alternationen bzw. Phonotagmen im Sprachkorpus von substanzieller Bedeutung. Ein Beispiel für eine morphologische interne Rekonstruktion wird mit der semantischen Methode geboten. Aus dem Vergleich von Wortpaaren, die in derselben semantischen Relation zueinanderstehen, können signifikante Phoneme oder Phonotagmen bestimmt werden, die auf ein unproduktives Derivationsaffix der *Vorsprache* hindeuten kann.

Acknowledgement

Writing a doctoral thesis can be an exhausting undertaking, but it is also a satisfying task for people who are interest in the subject of their dissertation. Such a project, however, is never the work of one person alone. This thesis is the result of my studies that was supported by many people whom I would like to thank for their inspiration, advice, and encouragement. First and foremost, I want to acknowledge my thesis supervisor Professor Gerhard Jäger, who welcomed me as PhD student in his research team and gave me enough freedom and room for conducting research in the field of my personal interest. I very much appreciate his fruitful advice, constructive criticism, and guidance which made it easy for me to keep going. I would also like to express my thanks to Professor Martin Kümmel wo agreed to take the role as co-supervisor. Ultimately, I want of thank my supervisors for their patience and understanding as my doctoral studies took longer than originally planned.

An important contribution to the thesis was made by the research group of EVOLAEMP and its subsequent seminars. Many thanks to all the members for inspiring presentations, insightful discussions, the friendly atmosphere and the possibility to present and discuss preliminary results of my studies. The meetings of the group provided a rewarding environment for innovative ideas and algorithms, even beyond the scope of this thesis. More particularly, I am very grateful to Johannes Dellert for his feedback and our interesting conversations. Furthermore, I should mention the persons who influenced and shaped me as a researcher before I came to Tübingen, such as Eva Tichy, Sebastian Padó, and Anette Frank. They aroused my interest in methods and algorithms of computational linguistics and Indo-European studies.

Finally, I want to express my gratitude to my family and all the other people from my social environment whose support was less related to content but even more important. Especially, I thank my parents for all their love and support to whom I dedicate this dissertation.

Abbreviations and Symbols

Abbreviations

A.B.	abbreviation of the author	MHG	Middle High German
acc.	accusative	NHG	New High German
act.	active	nom.	nominative
aor.	aorist	ntr.	neuter
App.	figure in the appendix	N, NV, ...	nasal consonant and sequences with nasal consonants
ASR	Ancestral State Reconstruction	OCS	Old Church Slavonic
C, CV, ...	consonant and sequences with consonants	OHG	Old High German
CHL	Computational Historical Linguistics	OIr.	Old Irish
CM	Comparative Method	perf.	perfect
cp.	compare	PIE	Proto-Indo-European
dat.	dative	pl.	plural
def.	definite wordform	PMI	Pointwise Mutual Information
ed.	edited	PR	phonological rule
fem.	feminine	pres.	present tense
Fig.	figure	prt.	preterite
Finn.	Finnish	ps.	person
fn.	footnote	pt.	point
fut.	future	pts.	points
FV	free variant	R, CRV, ...	sonorants and sequences with sonorants
gen.	genitive	SC	sound change

Germ.	Germanic	Sect.	section
Goth.	Gothic	sg	singular
H, HC, ...	laryngeals and sequences with laryngeals	Skt.	Sanskrit
HEA	hypothesis of etymological allomorphy	S-O-V	subject-object-verb
idf	inverse document frequency	subj.	subjunctive
impf.	imperfect	T, TR, ...	obstruents and sequences with obstruents
ind.	indicative	tf	term frequency
inf.	infinitive	TS	transcription error
IPA	International Phonetic Alphabet	V, VC, ...	vowel and sequences with vowels
IR	internal reconstruction	Ṽ	nasalized vowel
it.	Iteration	V2	verb-second word order
Lat.	Latin	VL	verb-final word order
MA	morphological alternation	vol.	volume
masc.	masculine	XSAMPA	Extended Speech Assessment Methods Phonetic Alphabet

Bibliographical Abbreviations

CHOIROBOS-KOS	Cramer, John Anthony (ed.): *Anecdota Graeca e codd. manuscriptis bibliothecarum Oxoniensium*, vol. 4. Oxford 1897.
CRATYLUS	Kratylos: De recta nominum ratione. Translated by Friedrich Daniel Ernst Schleiermacher. In: *Platons Werke*, series 2, vol. 2. 3rd edition. Berlin 1857.
DUDEN	Dudenredaktion (ed.): *Duden. Das Herkunftswörterbuch. Etymologie der deutschen Sprache* (= Der Duden in zwölf Bänden, vol. 7). 6th edition. Mannheim, Berlin, and Hamburg 2020.
KLUGE	Kluge, Friedrich / Seebold, Elmar (ed.): *Etymologisches Wörterbuch der Deutschen Sprache*. Revised 25th edition. Berlin and Boston 2011.
LIV	Rix, Helmut (ed.): *Lexikon der indogermanischen Verben. Die Wurzeln und ihre Primärstammbildungen*. Revised 2nd edition (with contributions by Martin Kümmel, Thomas Zehnder, Reiner Lipp, and Brigitte Schirmer). Wiesbaden 2001.
PĀṆINI	Vasu, Śrīśa Chandra (ed.): *The Ashṭādhyāyī of Pāṇini*, vol. 1. Translated by Śrīśa Chandra Vasu. 3rd edition. Allahabad 1896 [reprint Delhi 2003].

Symbols

*	denotes a reconstructed wordform (i.e., not preserved in any written documents)
†	denotes a fictive wordform
<	"comes from" or "is derived from"
>	"becomes regularly"
→	"turns into"
~	separates alternating sounds or wordforms
:	separates sound pairs/correspondences or word pairs
[ ]	IPA transcription
/ ... /	phonemic notation
⟨…⟩	graphemic notation

Spelling Conventions

All wordforms are written in italics. The meanings of linguistic forms are enclosed between single quotation marks. Semantic concepts are written in small capitals.

Content

1. Introduction

Present-day language is the result of a permanently ongoing process of change, and in this respect, change itself is reflected in the modern language. Speakers of languages may perceive the relics of past epochs as irregularities because they are contrary to the speaker's linguistic competence. The wordforms *was* and *were* cannot be explained synchronically and have resisted the tendency of regularize for a long time. Most linguistic change processes, on the other hand, are not apparent to the speaker, either because the process has left no trace in the language or because the relict does not synchronically violate any linguistic rule. The dominance of monosyllabic words in the English basic vocabulary is due to vowel loss, but this is as "inconspicuous" to the speaker as the word *barleycorn*, which indicates the older meaning of *corn*. The linguistic methodology of concluding the diachrony of a language from its synchronicity alone is subsumed under the term *internal reconstruction* (hereafter *IR*). The advantage of this method is obvious: without the additional knowledge of related languages or older language stages, statements can be made about historical stages.

Nevertheless, the dominant method in linguistics is the so-called Comparative Method. It compares words (i.e., *cognates*) in two or more related languages that go back to a common proto-form in the last common language stage and reconstructs the proto-form as well as the corresponding sound changes. Performing this method mechanically has been a central topic of *Computational Historical Linguistics* (CHL) for quite some time and has been able to make enormous progress during the last two decades. Its implementation is often associated with benefits in work with poorly studied languages but also with expectations in more detailed insights into the proto-languages.

This thesis, in contrast, is concerned with the automation of IR of ancient languages. The goals of IR are — as will be shown in this thesis — often different. This thesis deals with the theoretical as well as practical questions that are necessary to derive diachronic processes from purely synchronic data of only one language. The aim is less to mimic the linguistic process mechanically but rather to develop and test new computational methods. The first part comprises a theoretical treatise on IR as a scientific method and its importance for the subject of historical linguistics.

Understanding the aims and applications of this method is a prerequisite for understanding the path that automated IR should take in the future. To a certain extent, this part is also intended as a compendium on the subject of IR, which has not yet been presented in a comparable form. The second part deals with the automation possibilities of the methods. In general, I tried to keep the knowledge for the non-specialist reader as low as possible, especially for readers with a purely linguistic background, since they are the "end-users" of this very research-oriented topic. Means and methods are presented as far as they are necessary for understanding. Where an adequate presentation is not possible without digressing too much from the topic, reference must be made to the literature given.

The second and third chapters will serve as a basic presentation of the topic of linguistic reconstruction in general, and IR in particular. IR is only one of several methods of linguistic reconstruction and should therefore always be understood in the context of the alternatives in terms of their objectives and field of application. Fundamental questions and problems of reconstruction are described, on the one hand, to introduce non-historical linguists into the subject, and, on the other hand, to discuss the different views of linguists. In many cases, these different views are necessary to establish new internal reconstruction methods. An overview of the development of these views and the history of science will be given in Chapter 5. Chapter 4 will present an overview of the sub-methods of IR that have been discussed in the literature to date. The question which of these are suitable for automation will be discussed in the second part of the thesis. Some of them are inappropriate for automation due to the lack of available resources. Others are poorly researched, so their validity needs to be evaluated. As will be addressed in Chapter 6, the reliability of IR as a reconstruction method has repeatedly been criticized in the literature. Optimizing and validating the methods of IR will be an area of future work on automated IR.

The second part of thesis is the computational part of the thesis. The focus will be on the phonological reconstruction of sound change. For this purpose, two IR methods will be implemented and evaluated, each in three different variants. Besides the question of the ability to reconstruct, concrete questions of IR, which have been posed in the literature, will be answered as far as possible. German will be used as the evaluation language and Proto-Indo-European as the application case. In the last chapter, the work will be recapitulated, and the results of all methods are directly compared. By publishing the source code, resource files, and the tool *eo ipso*, a repetition for other languages should also be possible, provided that the data is available in a suitable form. The documentation and manual for *eo ipso* will be published separately and will be further developed at https://github.com/ambischoff/EoIpso. In this way, it should be possible to compare the results with those of other languages.

2. Reconstruction Methods in Historical Linguistics

August Schleicher is generally regarded as the founder of the scientific reconstruction of ancient languages. By comparing the preserved lexemes in successor languages, he attempted to restore original Indo-European wordforms (called proto-forms). In addition to the reconstruction of words, this also included to a certain extent the reconstruction of sounds and of one text: a short fable that he published by in 1868. It is precisely the reconstruction of texts that gives the reader — and not without good reason — the impression of a claim to linguistic-historical reality, but Schleicher (1869:342) stressed that the establishment of these proto-forms does not necessarily mean that these proto-forms had actually existed at some point in the past.

This tentativeness is also why Schleicher tried to distinguish the reconstructed proto-forms from attested forms with a preceding asterisk. In the subsequent period, a discussion was provoked in the academic world about what could be reconstructed at all and how these reconstructions were to be regarded (cf. Hermann 1907:2–16). Nevertheless, the question about the methods for reconstruction itself remained an uncommon issue in historical-comparative linguistics. Traditionally, most introductory works have illustrated a method by means of examples (the "pedagogical approach") and discussions usually only arose in polemical writings (cf. Hoenigswald 1950:357, Kay 1964:V). After 1870, the so-called Neogrammarians shaped a new perception of sound change and reconstruction with their positivist view of the "exceptionless of sound laws." They also began a more comprehensive treatment of sound change as a scientific subject. Some assumed that only by accepting the regularity of sound changes was reconstruction possible at all (cf. Hoenigswald 1950:364).

2.1 Reconstructed Forms

2.1.1 Are Reconstructed Forms Claimed to Be Historically Real?

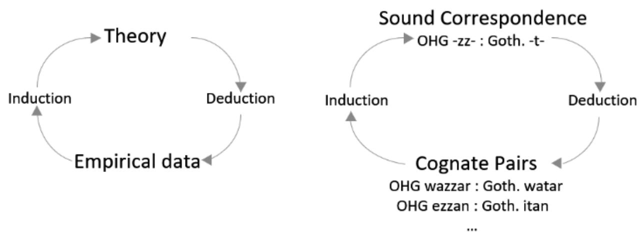

Figure 2.1: Illustration of induction and deduction as a part of the epistemological cycle of reasoning.

For non-linguists, it is difficult to assess the extent to which linguistic reconstructions have a claim to historical reality or whether they are to be regarded primarily as a purely theoretical construct. From an epistemological point of view, linguistic reconstructions can be described as theoretical products deductively inferred from an inductive theory. A reconstruction's truth value, thus, depends directly on the validity of an inductively reasoned theory. Inductive reasoning is inevitably based on different probabilities (cf. Givón 1999a:112), which indirectly influence the claim of reality for the deductively inferred reconstructions. Assume we are faced with the observation that all 33 Old High German words containing *-zz-* in the intervocalic position are cognates with Gothic words having *-t-* in the corresponding position. From this observation, we might induce a generalized theory that a Gothic *t* corresponds to the Old High German *zz* in the intervocalic position. Through this method of deductive reasoning, we could conclude that the Old High German *flazza* 'palm of hand' or 'sole of foot' corresponds to the Gothic **flato* even though this word is not attested in the Gothic texts as a whole (see Fig. 2.1). This sound correspondence is regular and relatively well attested. In contrast, the Old High German initial *fl-* corresponds to Gothic *fl-* or *þl-*, but there is insufficient evidence to determine this correspondence beyond a reasonable doubt (see Matzel 1962:220). Therefore, the reconstructed Goth. **flato* cannot be considered reliable from a methodological point of view, and its probability of its existence may ultimately depend on other (including extra-Germanic) evidence. In the end, the probability of any reconstructed wordform results from the probabilities of the sound changes that are taken into account.

However, even in the case of reconstructions based exclusively on regular sound laws, an absolute guarantee cannot be assumed. Further factors, which may be difficult to determine, may influence claims of historical authenticity. For example, a lexeme could be absent in the proto-language (a loanword or neologism), or there may have been an analogical or grammatical change that made the reconstruction of the correct proto-form infeasible. However, this uncertainty can be reduced by including further sister languages. The more cognate languages agree on these factors, the more likely it seems that these forms might have a proto-lingual origin. The possibility of late common proto-lingual developments that were more recent cannot be ruled out here either. In addition, there are cases where no daughter language has preserved a specific part of a proto-language's phonology or grammar. Notably, an absence of relicts is to be expected for irregular proto-forms that later tend to be regularized by daughter languages, as well as for cases in which the proto-language already provides the basis for later developments. For instance, a parallel sound deletion in all daughter languages is possible if a weak pronunciation of the proto-sound preceded that deletion.

Consequently, a linguistic reconstruction can only be interpreted as an ATTEMPT AT APPROXIMATION, behind which there is a model of inductive reasoning with varying degrees of probabilities. Thus, from a methodological view, it seems advisable to formulate these probabilities in concrete terms, but many of these factors may be difficult to grasp mathematically, such that an adequate model has not yet been presented.[1]

2.1.2 What Can Be Reconstructed?

Intuitively, it can be concluded that only what is preserved or otherwise reflected in daughter languages can be reconstructed. Only words preserved as loanwords in other languages could be excluded from this rule. In early stages of systematic and scientific reconstruction of ancient languages, the possibilities of linguistic reconstructions in general were debated: Which parts of proto-languages are "reconstructable" and how were they to be interpreted? Four different views emerged in this discourse.

[1] Attempts have already been made by the mathematician and linguist Hermann Graßmann (1860:22), but the number of providing examples varies, which makes it difficult to calculate any probability. Modern models, such as Hruschka et al. (2015), are mathematically more satisfying. However, these models depict rather sound correspondences than sound laws (see Sect. 7.1.2).

2.1.2.1 Only Words Can Be Reconstructed

Very early on, the idea of being able to completely reconstruct proto-languages encountered widespread criticism. The reconstruction of texts, in particular, implied that the proto-language is the sum of reconstructed forms. With respect to vocabulary that had potentially been lost, changes in meanings, and an inadequate ability to identify innovations, it was necessary to accept an irreconcilable divergence between reconstructed and historical vocabulary. Schmidt expressed it more provocatively:

> *Wenn wir also einen zusammenhängenden Satz in der Ursprache schreiben wollen, so kann es leicht geschehen, daß er, wenn auch jedes Element derselben richtig rekonstruiert ist, als Ganzes dennoch nicht besser dasteht als die Übersetzung eines Verses der Evangelien, deren einzelne Worte man teils aus Vulfilas, teils aus des sogenannten Tatians, teils aus Luthers Übersetzungen entnommen hätte, da alle geschichtliche Perspektive in der Ursprache noch fehlt. Die Ursprache bleibt demnach bis auf weiteres eine wissenschaftliche Fiktion.*[2]

Schmidt believed that only the reconstruction of individual proto-forms was possible but not the reconstruction of proto-languages. This view seemed to be a direct response to Schleicher's essay (1868), in which he wrote the aforementioned Indo-European fable about "the sheep and the horse," the first written text in a reconstructed proto-language. More recently, the difficulty of reconstructing complex grammatical objects has been attributed to the lack of arbitrariness of form and meaning (cf. Hoenigswald 1992:33). The Comparative Method relies primarily on this attribute (Rankin 2003:184), which has made it difficult to adopt this method for non-phonological subjects.

[2] Schmidt (1872:30). Translation: "If we want to write a coherent sentence in the proto-language, even if every part of it is properly reconstructed, as a whole, it would be no better than the translation of a verse of the Gospels, in which the individual words were taken partly from Wulfila's, partly from the so-called Tatian's, and partly from Luther's translations since the historical perspective would still be missing in the proto-language. The proto-language, therefore, remains a form of science fiction for the time being."

2.1.2.2 Only Sounds Can Be Reconstructed

Bremer (1894) had a more skeptical stance. According to him, while sounds can be reconstructed, the reconstruction of proto-forms are, in many cases, merely fictitious. He points to two possible anachronisms that may occur with the Comparative Method. The first anachronism arises when two reconstructed words, in fact, belong to different time periods. However, an anachronism can also arise within a word by assuming a proto-form that has the effect of two sound changes (or analogical formations), one of which may have occurred only after the wordform had already been changed by a third innovation (Bremer 1894:8). This is the case when three innovations (*a* : *b* : *c)* occurred, but the innovation *b* was not reconstructable. Bremer did not provide a concrete example, so his statement is illustrated in Fig. 2.2. In this example, the second change (an analogous formation with insertion of *e*) was not revealed. Consequently, the first step (an assimilation) could not have been detected, and finally, a wrong proto-form would be reconstructed by applying another sound change (*i* > *e*).

2.1.2.3 Nothing Can Be Reconstructed

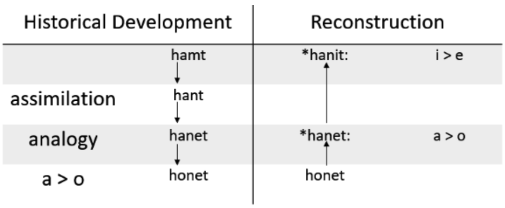

Figure 2.2: Fictitious example for Bremer's counterargument. The missing reconstruction of the step "assimilation" (left) causes a wrong proto-form (right).

The syntactician Berthold Delbrück, who devoted little of his own time to the reconstruction of Indo-European phonology and morphology, went one step further. He (1880:124) concluded that a proto-language is nothing but a formulaic expression for a scholar's changing views on the scope and nature of verbal material. It must be mentioned here that his view is caused by problems of the 19th century's linguistics. Delbrück interpreted the major changes and disunity on the reconstructions of the early Indo-European studies of his time as a potential counterargument for linguistic reconstructions per se. Later reconstructed proto-forms showed a much higher level of agreement. Nevertheless, some linguists (e.g., Pedersen 1905:398–399, Oertel 1897:416) maintained their position that reconstructions were to be regarded as mere "formulas." From this point of view, linguistic reconstructions are not admissible but rather only sound and word correspondences. Proponents of this view only considered equations without proto-forms (Goth. *h* = Lat. *k* = …) instead of sound laws like PIE **k* > Goth. *h*.

A very similar view can be found in the work of Meillet (1937:41), who asserted that an assumption of commonality of the corresponding languages can be concluded from correspondences, which would constitute the real object of historical linguistics. However, the potential common proto-language would remain completely inaccessible: "On ne restitue donc pas l'indo-européen" (Meillet 1937:41). Hermann (1907:6) already indicated the difficulty this view entails since it would be nearly impossible to carry out Delbrück's skepticism in practice. The interdependence of etymology and sound change means that the correctness of a sound correspondence depends on the correctness of the etymology and vice versa. This process implies some kind of reconstruction even if there are no suggested proto-forms.

2.1.2.4 Proto-languages Can Be Partly Reconstructed

More frequently, linguists have taken the more moderate position that reconstructability ultimately depends on each individual reconstruction. The Indo-European word for 'to be' is far more supported than the Indo-European word for 'table' due to the wealth of evidence for it in the daughter languages. This means that in linguistic reconstruction, other than a few valid examples, there are multitude of more uncertain instances, and the phonology may be too rich and too harmoniously structured to be natural (Hermann 1907:7). This "partial reconstructability" does not diminish the validity of the Comparative Method since proto-forms are to be understood as approximations to historical wordforms and have no immediate historical claim.

2.1.3 Are Reconstructed Sounds Claimed to Be Historically Real?

Hermann (1907:11) pursued a thought experiment on the reconstructability of proto-sounds. Suppose that only four Indo-European branches had survived. These languages could provide the following sound correspondences of the alveolar plosives, as shown in Tab. 2.A. Unless an explanation for the discrepancy of the Celtic-Balto-Slavic-Albanian *d* and the Germanic *t* and *ð* can be found, one would, at best, conclude that there was a Proto-Indo-European triconsonantism of **d* : **þ/t* : **ð* and the Germanic languages had split off very early. Otherwise, one would assume that there were only two alveolar plosives. Nothing here would indicate the "correct" reconstruction **d* : **t* : **dh*.

Table 2.A. Sound correspondences of alveolar plosives in Proto-Celtic, Proto-Germanic, Proto-Balto-Slavic, and Albanian.

Celtic	**Germanic**	**Balto-Slavic**	**Albanian**
d	t	d	d
t	Þ	t	t
d	ð	d	d

The PRINCIPLE OF APPROXIMATION is not limited to proto-forms but also applies to proto-sounds. As is commonly known, the exact pronunciation of a sound is by no means the same for all speakers of the same language, and even the same speaker may vary the pronunciation of a single sound (cf. Hirschfeld and Stock 2014:265–269 and Witting 1965). Additionally, a vowel preceding a nasal consonant and a vowel preceding another consonant may have a slightly other pronunciation, since the velum already begins to lift during the pronunciation of the vowel. The slightly nasalized vowel would not usually be perceived as an allophone. In further linguistic development, this sub-phonemic nasalization may be strengthened and may subsequently be established as an allophone. Due to loss of the nasal consonant, the nasalized vowel finally reaches a phonemic status. Since the proto-language's sounds can only be detected by sound change, sub-phonemic features can rarely be reconstructed (cf. Kurylowicz 1964:35), and suprasegmental features may not be reconstructed at all. Thus, in principle, linguistic reconstruction takes place at phonemic level.

This observation was already made in pre-structuralist years, when linguists were speaking of "l'archétype idéal" (Henry 1896:58) of a sound. It cannot be inferred from this, however, that proto-language's allophones are not reconstructable. The PIE allophone [z] of **s* (Tichy 2009:29) can only be reconstructed because both allophones are detectable by sound changes in daughter languages. Nevertheless, this does not definitively determine whether this **z* was pronounced as an alveolar, post-alveolar, or apical sound. The daughter languages refer to nothing more than a sound *similar to* [z]. In this context, the reconstructed sounds may only be understood as approximations, as well.

2.1.4 Interdependence between Reconstruction and Sound Change

Linguistic reconstructions are based on the application of detected sound laws. The principles for the extraction of sound laws and the interdependence of etymology and sound correspondences have likely been known since the beginning of Comparative Linguistics; however, they have rarely been discussed. Seebold described the interdependence of etymology and sound correspondences and the potential problems associated with them as follows:

> *Lautgesetze gewinnen wir also aus Entsprechungsregeln, und diese aus Wortgleichungen — und die Wortgleichungen ihrerseits bekommen einen ausreichenden Grad von Verbindlichkeit durch den Nachweis, daß sie durch regelmäßige Lautentsprechungen (oder historisch gedeutet: durch Lautgesetze) gestützt sind. Diese gegenseitige Abhängigkeit kann zu Zirkelschlüssen führen, aber nur solange, bis das Korpus der durch die angesetzten Lautentsprechungen gestützten Gleichungen so groß ist, daß die durchgehende Regelmäßigkeit der Entsprechungen nicht mehr dem Zufall zugeschrieben werden kann.*[3]

In other words, it should be noted that a sound law, which was initially hypothesized from word correspondences, can be disproved or confirmed by applying it to a larger corpus of "unseen" data (i.e., to new word correspondences). For linguistic studies, this entails some uncertainty with regard to weakly attested sound changes, especially when findings are contradictory or controversial. In the traditional model of sound development, a sound change is thought of as a temporally punctual event. An alternative perspective on sound change, however, was proposed by Wang

3 Seebold (1980:434). Translation: "thus obtain sound laws from [sound] correspondences, which are, in turn, derived from word correspondences; and word correspondences, on their part, can assume a sufficient degree of reliability by proving that they are supported by regular sound correspondences (or interpreted historically by sound laws). This interdependence may lead to circular reasoning, but only until the corpus of supporting correspondences is so large that the continuous regularity of the correspondences can no longer be attributed to chance."

(1969). In studying Chinese, Wang (1969:11) suggested considering "sound changes in coincident, incorporating, and overlapping relations with each other." Sound changes may modify themselves gradually over time by expanding or constricting to other sounds or environments (cf. Chen 1976:214–216). This thesis does not subscribe to such a view of sound change for pragmatic reasons, as it would complicate the interdependence model and may be difficult to conceptualize.

2.2 Methods of Linguistic Reconstruction

2.2.1 Reconstruction Method and Inferring Method

Formally, a linguist proceeds through three steps when conducting a linguistic reconstruction: First, word correspondences are defined as potential cognates in the PRESUPPOSITIONAL PHASE. This is followed by a construction phase, the actual RECONSTRUCTION METHOD (a method of inductive reasoning), and by a reduction phase, hereinafter called INFERRING METHODS ("principles of reconstruction" according to Milewski 1973:108–109). It is not uncommon for both the reconstruction and inferring methods to not be differentiated in scholarly literature even though they are based on different starting points (e.g., Bonfante 1945, Latta 1978). Any reconstruction method is ultimately based on one or more basic hypotheses and basic data. By applying the basic hypothesis to the basic data, both cognates (i.e., word correspondences) and sound correspondences are obtained by means of a method of inductive reasoning already described in Sect. 2.1.1. Through inferring methods, the sound correspondences are transformed to sound laws (*sound law reconstruction*), and cognate pairs to proto-forms (*proto-form reconstruction*). The reconstruction method usually entails a variety of *possible* proto-forms, so the primary task of the inferring methods is to reduce the number of possibilities.

According to Meillet (1937:41), sound correspondences are the only reality ("la seule réalité") of Comparative Linguistics. According to such an assessment, the final reconstructions derived through inferring methods have a purely hypothetical nature. Inferring methods are principles that attempt to exclude false reconstructions, which are frequently based on typological findings (Hoenigswald 1992:32). It is difficult to formally define inferring methods, since they often have no exclusive function. Their variety and validity are an issue in comparative linguistics. Divergent reconstructions by different linguists are often based on different inferring methods. Occasionally, TYPOLOGICAL PLAUSIBILITY in semantic, phonological, or syntactic terms is required for a final postulated diachronic change (see Givón

1999a:120). This approach should exclude “implausible” reconstructions by considering recurrences in other languages. However, the question of whether this actually identifies false sound changes is sometimes viewed critically. Lass (1980:75,90,92) argued that deductive-nomological explanations cannot be invoked in language change, which suitable rules (including inferring methods) do not, in principle, allow to be formulated (for counter-arguments and less critical views, see Harris 1995 and Keller 1994).

Various examples of inferring methods have been provided in the literature. Fox (1995:166) defined the distinctiveness of a reconstructed phoneme as the first principle and called for a “balance of the [reconstructed, A.B.] system and of typology.” In most cases, however, inferring methods are applied to sound correspondences. Examples of such methods are therefore given:

- PRINCIPLE OF ECONOMY (OCCAM’S RAZOR)
  „[W]hen there are two alternative hypotheses which attempt to account for the same set of data, the best motivated formulation of the sound changes is the more highly valued hypothesis “[4] (Latta 1978:35)
- PRINCIPLE OF PARALLEL EVOLUTION
  „[W]hen two linguistic elements, <u>x</u> and <u>y</u>, at one stage in the history of a language are reflected as <u>x'</u> and <u>y‘</u> respectively at a later stage, and if <u>y‘</u> bears the relation to <u>y</u> that <u>x‘</u> bears to <u>x</u>, the histories of <u>x‘</u> and <u>y‘</u> are assumed to be parallel.“ (Latta 1978:36)
- PRINCIPLE OF EXCLUSION OF UNCONDITIONED SPLITTING
  When a phoneme in language A corresponds to two or more phonemes of language B and it is impossible to determine any condition for sound change, then B has preserved the older state (cf. Milewski 1973:108).
- AREAL LINGUISTIC PRINCIPLES
  These methods try to derive “the age of elements from their distribution over a territory” (Milewski 1973:108–109). This can be done by (a) “peripheral archaisms,” which means that older elements are found in peripheral areas, or (b) the principle of “greater area,” which means that the older elements occur in the largest area.

[4] However, the unresolved question arises of what is to be understood as the “best motivated” or “simplest” solution. Latta (1978:128) refers to Halle (1964) for an attempt of such a phonological hierarchy.

Latta (1978:88) postulated inferring principles for the internal reconstruction method even if they can be applied to most reconstruction methods. In most cases, fewer inferring methods are accepted for IR due to the restriction to one language (e.g., Fox 1995:166). However, inferring methods for comparative reconstruction and IR have few differences other than their different basic data, as well as the fact that both methods work with sound correspondences as they were. Few principles are restricted to a special method (e.g., areal linguistic principles), so the distinction between "reconstruction method-specific" inferring methods is not made hereafter.

2.2.2 Basic Hypothesis

Linguistic reconstruction methods have been based on at least two basic hypotheses (cf. Latta 1978:28, "premises" according to Hajnal 2016:443). The first of these is the NEOGRAMMARIAN HYPOTHESIS, which stated that sound change affects all word-forms without exception. Sound change enables diachronic reconstructions and is therefore the basis of any reconstruction method. Although irregular changes are well documented, much data clearly speaks to the fundamental correctness of this hypothesis, such that Bloomfield (1925:130 fn. 1) regards it as linguistically universal. The regularity of morphophonemic alternations, likewise, supports the Neogrammarian hypothesis because synchronic alternations and diachronic sound change differ primarily only in temporal aspects. Alternative theses such as those of Wang (1969) and of Chen and Wang (1975), assumed a gradual spread of sound change from word to word, which was inferred from some observations in Chinese idioms. For a more detailed overview of this controversy, I point the reader to Labov (1981). In spite of all controversies, it seems that "until recently, the Neogrammarian had won the day" (Labov 1981:268).

The second thesis is the METHOD-DEPENDENT BASIC HYPOTHESIS and determines a method's subsequent approach. For the Comparative Method, this thesis refers to the relatedness of the postulated cognates (cf. Hajnal 2016:443, Givón 1999b:94). For IR, the hypothesis of etymological allomorphy is assumed meaning that all allomorphs date back to a unitary wordform.

In accordance with the definition by Peirce (1934:189), basic hypotheses as well as inferring methods are neither an inductive nor a deductive process but rather an abductive one. The term "abduction" describes an inference made by applying a result to a rule and another case (cf. Andersen 1973). In the already mentioned case of the OHG *-zz-* and the Gothic *-t-*, the Neogrammarian hypothesis is not apparent from the data but is rather a result of *another* inference. By applying basic hypotheses and inferring methods as abductive methods, the reconstruction process becomes an inductive method with abductive aspects (cf. Hajnal 2016:449). The different views on whether IR is an inductive (as Anttila 1973:349–350), deductive (Bonfante 1945:132, Fox 1995:181), or abductive (as Givón 1999b:91–94) method are ultimately rooted in the different concentrations on the sub-processes of IR, as well as their specific scope of application. The laryngeal theory, for example, works far more deductively in this respect than a simple internal reconstruction derived from a morphophonemic alternation.

2.2.3 Comparative Method

The methodology of linguistic reconstruction is now further exemplified through the Comparative Method. Its formal description largely follows the depiction of Hoenigswald (1950).

2.2.3.1 Cognate Recognition in the Comparative Method

The actual reconstruction process presupposes a set of cognate pairs as initial point. If a large proportion of these pairs is incorrect, this entails predominantly incorrect sound correspondences, and no further cognates can be recognized. The inductive cycle could not further progress, which is why cognate recognition is essential. With respect to the requisites for identifying two word as cognates, Katičić (1966:210) sets three conditions:

- The sound correspondences must correspond to the phoneme order.
- The sound correspondences must recur in other cognate pairs (also referred to as *recurrence*).
- The morphemes must be equivalent or similar.

There is no general agreement on how much sound correspondences must recur in data-sets. However, the principle that the greater the recurrence, the more reliably a sound correspondence can be articulated (Katičić 1966:210). In practice, the uncertainty as to whether a recurrence condition is actually fulfilled raises unresolved questions about Indo-European phonology. Nevertheless, *few* false correspondences can be identified in the further process and, thus, do not interfere with the cycle.

Table 2.B. Word and sound correspondences of Sanskrit and Germanic languages according to Hoenigswald (1950:358).

		Sanskrit	Germanic
(1)	t:t	*ásti* ‘is’	Goth. *ist* ‘is’
(2)	t:d	*pitár-* ‘father’	Goth. *fadar* ‘father’
(3)	t:Þ	*bhrā́tar-* ‘brother’	Goth. *broÞar* ‘brother’
(4)	d:d	*dehī́-* ‘wall’	Goth. *deigan* ‘knead’
(5)	d:t	*véda* ‘I know’	Goth. *wait* ‘I know’
(6)	dh:d	*mádhya-* ‘middle’	Goth. *midjis* ‘middle’
(7)	p:p	*spác-* ‘watcher’	OHG *spehōn* ‘look out’
(8)	p:b	*lip-* ‘smear’	Goth. *bi-leiban* ‘stay’
(9)	p:f	*pitár-* ‘father’	Goth. *fadar* ‘father’
(10)	b:b	*bódhati* ‘perceives’	Goth. *ana-biudan* ‘charge with, bid’
(11)	b:p	*rámbate* ‘hangs down’	MHG *lampen* ‘droop’
(12)	bh:b	*bhrā́tar-* ‘brother’	Goth. *broÞar* ‘brother’

2.2.3.2 Reconstruction Method of the Comparative Method

The initial point of the Comparative Method is the *basic hypothesis*, which assumes a genetic relationship of the languages to be compared. This hypothesis can be refuted or substantiated by an inductive reasoning process. It is not clearly stated what conditions a proto-form must have in order to be considered to be proto-lingual. Meillet (1925:340) assumed that at least three languages should preserve a word to assign it to the proto-language. On the other hand, areal linguists demand geographical conditions (cf. Sect. 2.2.1).

Word correspondences tend to suggest a list of sound correspondences. For example, Tab. 2.B shows word and sound correspondences for Sanskrit and Germanic languages according to Hoenigswald. In fact, the determination of sound correspondences from cognate pairs is a more formally complex process than is depicted by Hoenigswald (for more details, see Kay 1964:8). In the next step, these correspondences are classified by their sounds in both languages. The result of this step is sets that Hoenigswald (1950:358) called *partially alike sets*, illustrated in Fig. 2.3. These *partially alike sets* are sorted by “mutual exclusive environments,” which later emerge as the conditions of the sound change:

- d:d → occurs if the following syllable starts with Skt. *h* = Germ. *g*, *dh* = *d* and some more
- dh:d → does not occur in special environments
- t:d → occurs after unstressed vowel in Skt. and some other environments

Hoenigswald (1950:359) called the next step "fundamental assumption of comparative grammar," which transforms the sound correspondences into sound changes. He defines this process as follows: "PARTIALLY LIKE SETS OCCURRING IN MUTUALLY EXCLUSIVE ENVIRONMENTS ARE TAKEN TO BE CONTINUATIONS OF ONE AND THE SAME PHONEME OF THE PROTO-LANGUAGE [emphasis in original, A.B.]." With this assumption, we leave this comparative scheme of the reconstruction method, which until now has only been established on the basic hypothesis of a genetic relationship between the cognates, and move on to the inferring methods.

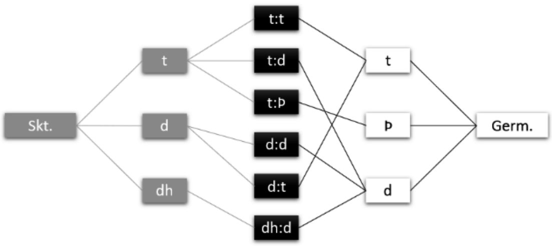

Figure 2.3: Sample of "partially alike sets" derived from the word correspondences of Tab. 2.B (Hoenigswald (1950:358).

2.2.3.3 Inferring Methods of the Comparative Method

If the previous method has been used to extract sound correspondences and their environments, the phase of reduction must follow. Without such a reduction phase, we would reconstruct twelve proto-sounds for each sound correspondence of Tab. 2.B. However, a sound correspondence does not necessarily have to originate from its own proto-sound. It is possible that a sound correspondence may be derived from two merged sounds (Fig. 2.4). The situation becomes more complicated in the case of multiple mergers, multiple phonemic splits, or both.

If we want to reduce the first six sound correspondences of Tab. 2.B as far as possible, while taking into account all *mutual exclusive environments*, there will be at least three proto-sounds and two possibilities (examples from Hoenigswald 1950:360):

- D_1 (= t:t, d:t), D_2 (= t:d, t:Þ) and D_3 (= d:d, dh:d)
- D_1 (= t:t, t:d, t:Þ), D_2 (= d:d, dh:d) and D_3 (= d:t)

Aiming for as few proto-sounds as possible, in this example, Hoenigswald applied the first inferring method: the principle of economics (2.2.1). Although this principle is justified, it must be emphasized that the principle is based merely on its own declaration. At this point, there are no other reasons to exclude possibilities with four proto-sounds.

The final decision between both possibilities depends more or less on the plausibility of a linguist's argumentation. From the first possibility, one would probably derive the proto-sounds D_1 = /d/, D_2 = /t/, and D_3 = /dh/ and attribute the first cognate pair *ásti : ist* to **-sd-*, which seems phonetically implausible. Therefore, Hoenigswald (1950:360) tends to reconstruct D_1 = /t/, D_2 = /dh/, and D_3 = /d/, applying the principle of phonetic plausibility as a second inferring method to further reduce the possibilities of reconstructions.

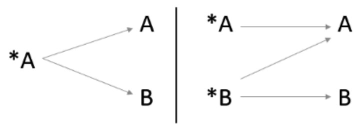

Figure 2.4: Simplified scheme to illustrate the relationship between proto-sound and sound correspondences. In the first case (left), the sound correspondence A:B goes back to one proto-sound; in the second case (right), the sound correspondence goes back to two different proto-sounds.

2.3 Types of Reconstruction Methods

Although the Comparative Method was able to establish itself as the standard method of linguistic reconstruction, new reconstruction methods have been repeatedly proposed and debated throughout history. Chapters on IR in the standard works and introductions to historical linguistics have been regularly found since the second half of the 20^{th} century (Lehmann 1962, Anttila 1972, Fox 1995). The other methods that are mentioned and addressed frequently reflect the author's research focus and *zeitgeist*. Hockett (1958:461) cited four linguistic reconstruction methods: external method (i.e., Comparative Method), IR, dialect geography, and glottochronology, although dialect geography is usually no longer understood as its own reconstruction method. Birnbaum (1977:17), on the other hand, rejected glottochronology as a method and added reconstruction via loanwords as a third variant (Birnbaum 1970:97). Greenberg (1978:70) observed four methods by additionally listing the "historical method" (corresponding to the Philological Method).

In more recent introductions, a restriction to the Comparative Method and, if any, to the Internal Reconstruction seems to have prevailed, with both methods being regarded as standing "in a complementary relationship to one another" (Unger 2000:655). In this two-method-attitude, the Comparative Method is considered "the most reliable, and the most legitimate, of the available methods" (Fox 1995:145).

The "three methods of reconstruction" according to Chen (1976) are to be distinguished from reconstruction methods. By those he meant methods of relative chronologies. He used the term "external method" for the Philological Method, while the term "latitudinal method" signified the application of the Comparative Method to relative chronology (Chen 1976:210, 237–239).

2.3.1 Reconstruction Methods According to Bonfante

Typologies listing more than four reconstruction methods are often based on the work of the areal linguist Giuliano Bonfante. In his frequently cited paper, he (1945:144) defines ten different methods of linguistic reconstruction:

1. The method of the impossibility of spontaneous scission
2. The method of linguistic geography
3. The method of areal linguistics
4. The method of the general linguistic trend
5. The method of Internal Reconstruction
6. The method of the anomalous form
7. The method of the usual phonemic change
8. The method of linguistic paleontology
9. The method of the *fase sparita*
10. The method of the chronology of the texts

The first three methods are part of the Comparative Method, whereas the other methods "may be applied also to one language alone" (Bonfante 1945:144). The geographical method tries to draw conclusions on the older course from the present course of language boundaries or isoglosses (see Milewski 1973:108). This geographical method, as well as the points (2–4), (7), and (9), are actually inferring methods (for more details on these methods, see Sect. 4.3). The method of linguistic paleontology (8) adds (pre)historical knowledge as an extra-linguistic source to the reconstruction process (e.g., timing or spreading of animal domestications). Once again, this is an inferring method that can decide on the question of the older semantics of a word. However, it cannot reconstruct proto-forms on its own. Method (10) refers to the traditional philological approach, which compares texts of a single language from different eras. It can be used, for example, to determine sound changes between periods A and B and thus to reconstruct unattested wordforms of the period A.

The remaining methods (5) and (6), on the other hand, work in a "strictly synchronic" fashion (Bonfante 1945:144), but Bonfante did not explicitly distinguish between their procedures. He attributed the method of the anomalous form to Meillet and likely referred to Meillet's (1937) statement:

> *Si l'on ne possédait pas le latin et si les dialectes italiques étaient représentés seulement par le français qui n'a plus l'aspect général d'une langue indo-européenne, il ne serait pas pour cela impossible de démontrer, par de détails précis, que le français est indo-européen. La meilleure preuve serait fournie par la flexion du présent du verbe 'être'*[5]

Meillet referred, here, to the anomalous conjugation of the verb 'to be' in most Indo-European languages: French *(il) est : (ils) sont*, Skt. *ásti : sánti* or Goth. *ist : sind*. This parallelism can be attributed neither to coincidence nor to borrowing and indicates a genetic relationship between these languages. Nevertheless, Meillet did not see this comparison as a method of reconstruction but merely as a means of proving linguistic kinship. The distinction between both methods — Internal Reconstruction and the "Method of the anomalous forms" — was therefore made by Bonfante himself. Moreover, he oversimplified the issue with his thesis that analogy "levels paradigms and destroys 'anomalous' forms, it never creates them" (Bonfante 1945:133). This overlooks cases of spreading analogy, such as the spreading of the irregular Norse inflection *sá* 'to sow' : *sera* (< **se-zō*) 'sowed' to *gnúa* 'to rub' :

[5] Meillet (1937:37). Translation: "one had not possessed Latin and if the Italic dialects had been represented only by French 'that no longer' has the general appearance of an Indo-European language, it would not have been impossible to demonstrate, in precise details, that French is an Indo-European language. The best proof would be provided by declining the present of the verb 'être' ['to be', a.b]."

gnera 'rubbed'. In this example, the method would erroneously take the inflection of *róa* as an inherited one. Bonfante, however, seems to regard the method as an inferring method and, as such, as legitimate: the irregular form out of the two competing reconstructions is the older one.

It should therefore be noted that of the ten proposed reconstruction methods, seven should, in fact, be regarded as inferring methods. The lack of distinction between reconstruction methods and inferring methods results in a large number of methods that do not, in fact, reconstruct but can be applied to reconstruction methods. A similar conclusion may have been made by Milewski (1973:102–103), whose typology of methods is obviously based on Bonfante's classification. He grouped them under three methods: Philological Method, Internal Reconstruction, and Comparative Method. Milewski (1973:108–109) identified Bonfante's methods (1–4) and (7) as inferring methods (in his terms, "principles of reconstructions").

2.3.2 Interrelations between Reconstruction Methods

In this work, four fundamental reconstruction methods are assumed: The Comparative Method, Internal Reconstruction, the Philological Method, and the approach using loanwords, which will be referred to as the "External Method" hereafter. This term was coined by Birnbaum (1970:97) to define a reconstruction method "based on extraneous linguistic elements (borrowings, loan and foreign words, non-native proper names, etc.)." In his later works, the term also included "extra-linguistic evidence," which cannot be applied by itself but as a "supplementing" method for IR or Comparative Method (Birnbaum 1977:1[6]). Since this work is limited to linguistic reconstruction methods, extra-linguistic factors are not considered. However, the term "External Reconstruction" is questionable because Hockett (1958:461) and Mayrhofer (1982:177) equate this term with the Comparative Method.

Each reconstruction method is defined by its own BASIC HYPOTHESIS, which postulates a common origin of two wordforms and morphemes respectively and thus lays the foundation for proto-word reconstruction:

- COMPARATIVE METHOD
  The words of two different languages are inherited continuations (i.e., *cognates*) of the same word (*proto-form*) or root (etymon) of the common *proto-language*

[6] On another page in his book, Birnbaum (1977:60) used the term for another method.

- INTERNAL RECONSTRUCTION
  The morphemes of two wordforms of a single language are inherited continuations (i.e., *internal cognates*) of the same morpheme (*pre-form*) of an earlier stage of the language (*pre-language;* rarely called *palaeo-language*, Dolgopolsky 1989:13)
- EXTERNAL RECONSTRUCTION
  One of two words of two different languages (referred to as *donor language* and *recipient language*, respectively) is the adopted form (*loanword*) of the other or its pre-form (i.e., the original word of the donor language).
- PHILOLOGICAL METHOD
  One of two words is the historical ancestor of the other.

An important distinctive feature of each of these methods is its result. The Comparative Method reconstructs a PROTO-LANGUAGE, which attempts to reflect the state of the language at the time of splitting. In the case of external reconstruction, the state at the DATE OF BORROWING is restored for the language with the loanword. Internal Reconstruction cannot indicate anything about the time of its reconstructions, except that these reconstructions had existed in the PRE-LANGUAGE. Since External and Internal Reconstruction only reconstruct single words or sound changes, their reconstructions may belong to different time stages.

All four methods additionally presuppose the Neogrammarian hypothesis, which can be regarded as the "general" hypothesis of linguistic reconstruction. This hypothesis enables the falsification of reconstructions (Tichy 2009:23). In this work, three dualisms are defined for the four methods, which can further differentiate the single methods from each other (Fig. 2.5).

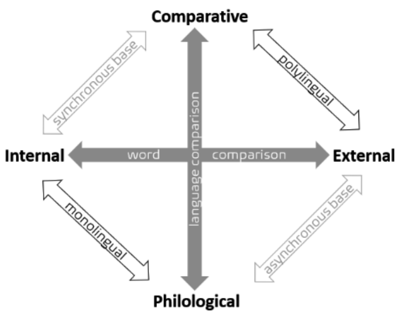

Figure 2.5: Display of reconstruction methods and their relation to each other. The three dualisms are marked in different colors.

Language versus word comparison | The first dualism differentiates between two types: methods that additionally extend their basic hypothesis to the language system and methods that leave their comparison at the morphological level. For the Comparative and Philological Method, not only can two morphemes be traced back to one origin (or there is a temporal relationship between them), but also both languages can be traced back to an earlier stage of the language. Only by this extension of the basic hypothesis can a PHYLOGENETIC RELATION between the languages be postulated. This distinction plays an important role in particular between Comparative and External Reconstruction. Since both methods work polylingually, it is often not possible to say for sure — even in the case of proven kinship — whether two morphemes are cognates or (early) loanwords. Additionally, the lack of phylogenetic relations in External Reconstruction also leads to the situation that the phonetic difference between both words is caused not only by sound change but also what is called *phonetic integration* (Eisenberg 2011:172). For example, the English words *baby* and *cakes* were adopted into German as /ˈbeːbi/ and /keːks/ because the diphthong *eɪ* is absent in German. Thus, a German sound change *eɪ* > *e* cannot be inferred from this.

Multilingualism | Of all the methods, IR stands out because it actually works with only one langue (synchronic monolingual). However, the philological method postulates such monolingualism by declaring both languages to be the same at different temporal stages (diachronic monolingual). These methods differ from the External and Comparative Method, which work with different languages (polylingual).

Synchronous vs. asynchronous relations | The term “asynchronous relation” between two comparative morphemes A and B is understood in this work as the additional assumption that morpheme A descends from morpheme B and thus belongs to a different time stage. In contrast, the term “synchronous” refers to the same temporal stage of reconstruction, which does not need to coincide with temporal synchronicity. According to this definition, the comparison of English and Latin is “synchronous” according to the Comparative Method even though both languages were not spoken at the same time. The diachronic aspect only enters the procedure through diachronic projection, which refers to the backward projecting of the reconstructed wordform to an earlier time stage (a similar model is provided by Fox 1995:207).

In methods with an asynchronous relation, a diachronic projection also takes place, but here, one of the compared wordforms itself is projected backwards. The External Method can postulate an additional wordform as a borrowed word, which is to be set temporally before or after the morpheme of the donor language (see Fig. 2.6). This is the case for Gothic *mes* 'table', which is not directly derived from the Latin *mēnsa* but from the Vulgar Latin **mēsa* (Sonderegger 1979:82). To illustrate this development, two projections are required: while the first one projects the Latin word backwards (dashed line in Fig. 2.6), the second projection places a reconstructed wordform *mesa* between Latin *mēnsa* and Goth. *mes*.

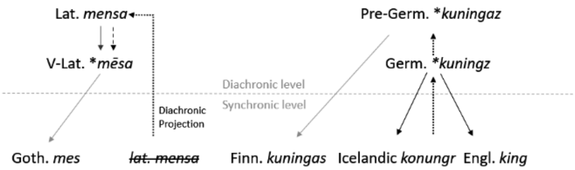

Figure 2.6: Examples of three-stage models with two projections. On the left, the second projection is postulated for a younger stage than mensa. On the right, both projections go back in time. For the reconstruction of **kuningaz* see Stanislav (1978:312). Grey line represents borrowing, black lines stand for kinship.

Reconstructions that assign the second projection to a further time depth are also possible (see Fig. 2.6). However, the question of whether External Reconstruction can actually produce double projections by itself must be considered. The Germanic reconstruction **kuningaz* cannot be derived without Comparative Reconstruction, and without knowledge of the Romance languages, it cannot be decided whether the nasal loss in *mēnsa* took place in Romance or Gothic language (or there was a secondary *n*-insertion in *mēnsa*). The external reconstruction itself provides only one diachronic projection. Additionally, inferring methods can provide answers for such questions. For example, if Gothic shows an *n* in comparable positions in non-loanwords, this may be an argument against a Gothic sound change. Conversely, an absent nasal insertion in other Latin words may be an argument against a secondary *n*-insertion in Latin. Through these considerations, even without comparative knowledge, one may arrive at the correct reconstruction.

2.3.3 Mixed Reconstructions

As the example of **kuningaz* shows, linguistic reconstructions, in practice, are mostly the result of hybrids of various reconstruction methods. Priority is given to the Comparative and Philological Methods, while other methods merely have a supplementary function.[7] However, a strictly individualized consideration of all methods — as done in this work — allows them to be evaluated in an analytical way.

The integration of further methods into a procedure relates to additional basic hypotheses, which signifies a corresponding adjustment of the probability of the final reconstruction. In contrast to reconstructions that lead to the same results through two independently performed methods and thus support each other, the probability of a particular result does not (necessarily) increase by adding further basic hypotheses. Examples of such mixed methods for IR are discussed in greater detail in this section. In these methods, we can observe initial data that are identical to that of IR, and the probability of a particular reconstruction decreases with each additional hypothesis. In practice, these additional hypotheses are therefore used to reduce otherwise unexplainable alternations. Functionally, the basic hypotheses of other methods are consequently applied as inferring methods.

2.3.3.1 Comparative-Internal Method

By combining Comparative and Internal Reconstruction, an external origin from a (closely) related language is postulated for one of the "internal cognates." The basic hypothesis of *internal* kinship is thus supplemented by an additional basic hypothesis of *external* kinship. With this additional assumption, on the one hand, an external relationship is postulated based on the internal without having to use this language for comparison. On the other hand, we must assume a proto-language rather than a pre-language. Proto-forms reconstructed in this way have the considerable weakness that, in contrast to a comparatively reconstructed proto-form, only one sound change of a proto-form can commonly be reconstructed since all later sound changes are carried out by a loanword (see the example of German *Waffe* in Fig. 2.7).

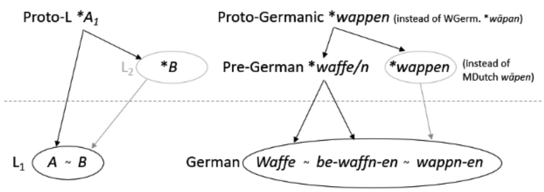

Figure 2.7: Illustration of the Comparative-Internal Reconstruction exemplified by the German words *Waffe* 'weapon' – *bewaffnen* 'to arm' – *wappnen* 'to brace oneself' (cf. Paraschkewow 2004:376).

[7] For example, Hajnal (2016:450) uses IR in his paper rather as an inferring method for the Comparative Method.

2.3.3.2 External-Internal Method

There are two options for applying an External-Internal Method: (1) The alternation can be traced back to a pre-form of the recipient language or (2) the alternation itself was borrowed from the donor language. The first case of applying an External-Internal Method occurs with *reborrowing*[8] and follows a scheme similar to that of the Comparative-Internal Method, without assuming any linguistic kinship (see Fig. 2.8 first depiction). Instead of proto-languages, pre-languages are reconstructed in this way. However, it is barely possible to distinguish reborrowing from a random suppletion through internal evidence.

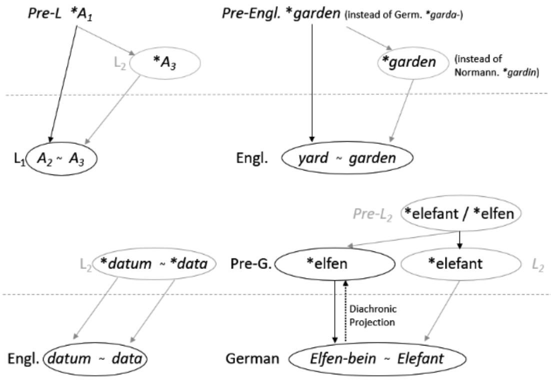

Figure 2.8: Illustration of two types of the External-Internal Method. The first type occurs with reborrowing and is exemplified by the English words *yard* : *garden* (Obst and Schleburg 2004:310). The second depiction illustrates the External-Internal Method with borrowed alternations. On the left, the synchronous type is illustrated by the English wordforms *datum* : *data*. On the right, the diachronic type is exemplified by the German words *Elfenbein* ‘ivory’ : *Elefant* ‘elephant’.

In the second case of applying External-Internal Methods, both words are considered to be loanwords. The alternation may have been adopted directly from the donor language (*synchronous type*) or may have entered the recipient language via two phases of borrowing (*diachronic type*). For both types, it should be noted that neither a pre-language nor a proto-language has been reconstructed for the recipient language. In the case of diachronic types, a diachronic projection takes place, projecting the older morpheme back to the time of the borrowing phase.

---

[8] This definition of *reborrowing* is in accord with Schultz (2012:52). Other linguists (e.g., Vitali 2007:283) use the term for borrowings from earlier language stages.

2.3.3.2 Philological-Internal Method

A Philological-Internal Method is applied when a historical borrowing is assumed for an alternation. Historical borrowings are revived words from past eras and presupposes the preservation of texts from earlier language stages. The obvious advantage in these cases arises from a “direct” diachronic projection: As in the Philological Method, sound changes can be inferred directly from the antecedent wordform. In the case of an “indirect” diachronic projection, the time of revival is also projected backwards, as secondary sound changes may have occurred. In addition to sound changes, *phonetic integrations* are to be assumed as they were in the External Method. The French example *préhension* in Fig. 2.9 illustrates this fact by comparing the pronunciations of the French and the Latin forms of the word.

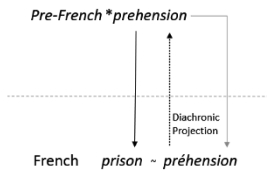

Figure 2.9: Illustration of the Philological-Internal Method exemplified by the French words *prison* ‘prison’ and *préhension* ‘prehension’.

3. Theory of Internal Reconstruction

3.1 What is Internal Reconstruction?

3.1.1 Definition

The attempt to find a uniform and satisfactory definition for Internal Reconstruction, including all the methods proposed in literature, is not a simple undertaking. The definitions put forward can partly contradict each other and have only in common that they are "non-comparative."[9] However, among the linguists' works two simplified definitions stand out.[10]

3.1.1.1 Synchronic-Monolingual Definition

One is the SYNCHRONIC-MONOLINGUAL DEFINITION, which can also be determined as the older one:

- PISANI (1938:32): "… uno, che io chiamo di 'ricostruzione interna', in cui los studioso, ammaestrato beninteso da ciò che gli insegnano le altre lingue dello stesso gruppo, trae dal materiale monoglottico tramandato (compreso quanto offre la ricerca degli antichi imprestiti ecc.) tutte le conseguenze possibili per raggiungere fasi anteriori,"[11]

[9] An idiosyncratic example is provided by Jeffers (1976:8), who denotes any form of integration of universals into (comparative) reconstructions as a kind of IR. The only commonality to other definitions is that universals were not inferred by the Comparative Method.

[10] Another typology of definitions can be found in Tremblay (2005:104-119), who distinguishes four meanings of IR: (1) by dialect comparison and loanwords, (2) the morphophonological method, which corresponds to the second definition, (3) by simple diachronic projection of a wordform and (4) the typological method, which refers to the term "external means." Tremblay's typology focuses on the different source data and supplementing possibilities rather than the different "definitions" (see also the commentary in Hock 2013:4).

[11] Translation: "One [method, a.b.], which I call 'Internal Reconstruction', in which the scholar, trained, of course, by what the other languages of the same group teach him, learns from the monoglot material handed down (including what the research of ancient material offers, etc.) all the possible consequences to reach earlier phases."

- BONFANTE (1945:85): "this method is not necessarily based on the comparison of one language with another (or of one dialect with another, which is the same thing); it can be applied on the basis of the observation of the system of one given language."
- CHAFE (1959:478): "Internal reconstruction is a procedure for inferring part of the history of a language from material available for a synchronic description of the language, and from that alone."
- BORETZKY (1975:49–50): "Allgemein definiert man IR als eine Rekonstruktion mit den inneren Mitteln einer Sprache, d.h. ohne Zuhilfenahme anderer Sprachen oder Dialekte, die eigene, von der zu untersuchenden Sprache deutlich unterscheidbare Systeme darstellen."[12]
- FOX (1995:146): "The essence of the method of Internal Reconstruction is that evidence for an earlier stage of a language can be deduced from certain *internal* patterns of the language, without recourse to comparative evidence from related languages. [original emphasis, A.B.]"

These definitions have in common that they demand a reconstruction in accordance with the singularity principle (i.e., on the basis of only one language), which is to be temporally and systematically delimited from other languages. With respect to this point, IR is opposed to the comparative principle of historical-comparative linguistics and thus differs from the Comparative Method in the number of languages that serve as the starting point for reconstructions. Going beyond the question of the initial data, Boretzky's definition tries to include the question about the approach but without explaining what does he mean by "inner means."

3.1.1.1 Structuralist Definition

The second definition for IR bases on this question and can be called STRUCTURALIST DEFINITION. It defines IR via the basic hypothesis itself:

- HOENIGSWALD (1965:132): "Internal reconstruction is based on the principle that phonemes which alternate represent, wholly or in part, former co-allophones."

[12] Translation: "In general, IR is defined as reconstruction with inner means of one language (i.e., without the use of other languages or dialects), which are separate systems of their own from the language to be examined."

- ROSÉN (in Kurylowicz 1964:33): "Coming back to what was traditionally termed, 'internal reconstruction', namely the rediscovery of lost phonemic or morphemic shapes of linguistic forms."
- LATTA (1978:1): "In its most straightforward form the method assumes that paradigmatic allomorphy is not original, but has been introduced by regular sound change."
- CAMPBELL AND GRONDONA (2007:5): "[A]n assumption underlying internal reconstruction is that the variants (allomorphs) of a morpheme all stem from a single invariant original form and that the variants are the result of conditioned changes that the language has undergone in its past."
- BYNON (1977:90): "The reconstruction of diachronic phonological rules on the sole evidence of synchronic morphological alternation is known as *internal reconstruction.* Although it is by no means an invention of structuralist linguistics […] only the techniques of descriptive linguistics allow the explicit statement of its underlying principles. [original emphasis, A.B.]"

These definitions draw a link to structuralism and descriptive linguistics (cf. Bynon 1977:90 and Sect. 5.5). As structuralism tries to map the multiplicity of variants (i.e., allophones, allomorphs) to an entity, so IR tries to map the alternation of variants to a historical entity. This definition does not violate the singularity principle, but according to this definition, IR is constrained in its methodological possibilities. Even though the application of linguistic results from empirical or typological research for linguistic reconstructions complies with the singularity principle, this would not be in accordance with the structuralist definition. Such information is derived from linguistic studies and is therefore referred to as "external means" in this thesis. "Internal means" are methods that do not go beyond the knowledge of linguistic competence, which would include morphophonemic rules or derivational formation patterns.

Depending on whether or not external means are included, it may make sense to distinguish between an IR in the *narrow sense* and an IR "in the *broad sense*" (Kurylowicz 1964:10). To Kurylowicz (1964:30–31), external means, such as universals, play an important role in IR. However, he excludes extra-linguistic knowledge (e.g., geographical or historical knowledge) from this definition:

> *The stress may be put on the strictly linguistic aspect excluding approaches of a mixed character, like e.g. areal linguistics […] Hence internal reconstruction also renounces collaboration with auxiliary sciences like experimental phonetics, psychology, or cultural anthropology. […] the difference [to the comparative method, A.B.] will lie in the conceptual equipment at our disposal, in our deepened insight into the nature and functioning of language. Reconstruction has also become more "internal" in the sense that in our reasonings we are more apt to discard considerations of experimental phonetics or of psychology — both of them favourite implements of reconstruction in earlier periods.*[13]

Kurylowicz goes one step further, by excluding all non-linguistic information, making the "pure" linguistic reconstruction method:

> *internal reconstruction means purely linguistic reconstruction, to be distinguished from other methods aiming at non-linguistic problems or at any rate at problems which are not purely linguistics, like the primitive home of the I.E. [i.e., Indo-Europeans, A.B.], the migrations of I.E. tribes, their culture, and so on. The status of linguistics as an autonomous science, independent of experimental phonetics, psychology, cultural anthropology, and so on, postulated by de Saussure, has become a fact not only in descriptive but also in historical linguistics.*[13]

The integration of extra-linguistic knowledge is thus the third way to conduct IR. Kurylowicz's definitions of "narrow" and "broad" IR are now mostly regarded as an "unrepresentative" alternative view.[14] The reason for this is that the inclusion of

[13] Kurylowicz (1964:30)
[14] This becomes clear in Fox (1995:182 fn. 1). Fox considers language typology being part of "comparative linguistics," while "internal to language" includes "comparative linguistic evidence."

typological means is understood as a violation of the non-comparative nature of IR. Nevertheless, this thesis includes both internal and external means as admissible possibilities to optimize the procedure of reconstruction.

A separate definition for IR is provided by Latta (1978), who has published the first doctoral dissertation that comprehensively addressed IR and concluded that IR is to be understood as six methodological principles that "jointly constitute the method of internal reconstruction" (1978:126). These principles include the basic hypothesis (called "principle of automaticity"), some inferring methods (see Sect. 2.2.1), and assumptions regarding the presuppositional phase ("principle of similarity of function"). Defining the methods of IR as a sum of principles does not seem appropriate to me since these principles are not to be regarded as factors that distinguish IR from other reconstruction methods. Instead, they appear in this definition as a disordered set of principles without discussing their actual function within the procedure.

3.1.2 The Term "Internal"

In his 1907 paper, Hermann called for a so-called single-language reconstruction ("Einzelsprachen-Rekonstruktion" 1907:63) before any comparative reconstruction takes place, which is often regarded as the first methodological description of IR (Szulc 1987:18). His method exemplifies a major difference with today's IR in that it excluded the concept of synchronicity: in Hermann's reconstruction process, both earlier language stages and dialects could be included (see Sect. 5.3). Since his method was conceived as a 'preparatory work' for language comparison, the word 'single language' referred to the languages used for comparison, which in practice could also be language families or sub-families.

In later times, however, the terms that prevailed tried to express a remaining "within" the language and emphasized its monolingual nature: internal analysis (Bàrtoli 1935:417), intrinsic evidence (Pagliaro 1930:174), internal reconstruction (Pisani 1938:32), or inner reconstruction (Boretzky 1975). According to Kurylowicz, the motive for identify the limitations of reconstruction apply not only within one language but also within the linguistic possibilities: "The internal character of reconstruction consists in avoiding any recourse to external, extralinguistic facts, including articulatory facts." (Kurylowicz 1964:12, see also Sect. 3.1.1.1).

3.1.3 The Term "Synchronic"

In addition, the term "synchronic method" (Fowkes 1950:142) is also used for the IR, which, in contrast to the monolingual nature of the term "internal," highlights the synchronic character of the IR. The synchronic-diachronic dichotomy distinguishes between a consideration of language that is limited to a particular time and another that takes into account the aspect of time. While ancient languages can also be analyzed synchronically, a strict separation of both synchronic and diachronic perspectives is not always possible or expedient (cf. Janda and Joseph 2003:121–123).

By its definition, IR minimizes other language stages and reconstructs an older state from only one — synchronically considered — language stage. It only creates diachrony through diachronic projection. This diachronic aspect is added by the basic hypothesis of etymological allomorphy with its diachronic premises and historical claim (see Sect. 2.3.2). The claim to historical reality or approximation is the basis for considering a linguistic method a "reconstruction method." This becomes all the more obvious when comparing this method with similar methods that do not claim to conduct reconstruction, such as phoneme analysis. The phonetic alternation of German [hʊn**t**] 'dog' and [hʊn**d**ə] 'dogs' is attributed to the abstract phoneme /d/ and the phonological rule d > t / _#. This rule works synchronically and has no historical claim, although a historical claim might be possible.

It is, therefore, not surprising that the principle of synchronicity has been called into question (see Sect. 5.6.2). According to Birnbaum (1970:106–108), the postulates, such as the assumption of a common pre-phoneme or the formulation of phonological rules, basically lead to the abandonment of synchronicity. These postulates can ultimately be assumed "either by certain linguistic universals or by our knowledge of some specific earlier processes in the evolution of that language" (Birnbaum 1970:106) and therefore — in an indirect way — work diachronically.

Synchronicity can therefore only be assumed for the initial data. Comparative Reconstruction works synchronically in the same way and generates diachrony by the hypothesis of genetic relationship. It does not matter whether the compared languages existed at the same time; both are considered synchronically. Notwithstanding, linguistic competence also implies an inherent diachronic knowledge. This includes, for example, archaisms or, more generally, knowledge about the productivity or unproductivity of inflectional patterns. This "handed-down" diachrony can be a starting point for IR but at the same time makes the distinction between the value of synchronic and diachronic studies. Kurylowicz (1964:9) clarified this succinctly:

"Everywhere a complete 'synchronic' description of a language must have recourse to the notions of archaisms and innovation."

3.1.4 Basic Hypothesis

The basic hypothesis was already found in Hermann (1907) but only claimed a central role in this reconstruction method because of the papers of Hoenigswald (1944, 1946, 1965). Marchand (1956:246) defines the basic hypothesis using an inherently circular argument: "It may be said that all the allomorphs of a morpheme are cognate with one another, unless suppletion has occurred." Knowing whether there is a suppletive paradigm cannot be determined synchronically. IR cannot offer more than the "hypothesis" that there was a common origin of the allomorphs (*hypothesis of etymological allomorphy*, HEA). It is well known that this hypothesis is by no means universally valid; in fact, it has been repeatedly criticized (see Sect. 6.2). Nevertheless, it has a sufficient probability in practice and also enables reconstructions of sound laws because the "interplay of sound change and analogy may create patterns so typical as to make it possible to recover from them the process to which they owe their existence" (Hoenigswald 1946:138). The basic hypothesis is therefore the solidest point of IR but also where it breaks down. It cannot provide more than a simplified and incomplete picture of the previous language stage, which has made it subject to a number of criticisms (see Sect. 6.2.3).

The majority of linguists conspicuously limit their definition of the basic hypothesis to the phonetic level, which also offers major implications for practical applications. Supporters of the synchronic-monolingual definition, on the other hand, have sought to define the basic hypothesis in more general terms:

> *The fundamental assumption [of IR, A.B.] is that some events in the history of a language leave discernible traces in its design, so that by finding these traces one can draw inferences as to the earlier incidents which are responsible for them.*[15]

The question of whether IR is based on further auxiliary assumptions in addition to the HEA can be answered differently. Ringe (2003) developed a second assumption

[15] Hockett (1958:463)

of IR, at least in a few cases, referring to cases such as the English nominal paradigms ending with *-f*, which are pluralized with either *-fs* (e.g., *oaf*) or with *-vz* (e.g., *loaf*):

> *To determine which paradigms are (in that sense) "old," we invoke a second assumption: paradigms which are irregular in terms of a language's current grammar are likely to be inherited, reflecting the regular grammar of an earlier period.*[16]

In principle, this rule applies to all non-automatic alternations (e.g., the *ablaut* paradigms in Germanic languages) and is not limited to paradigms (cf. Meillet's rule in Sect. 2.3.1). However, I may doubt that there is indeed a new hypothesis here. The HEA does not assume a diachronic projection for the non-alternating (i.e., regular) paradigms, so they are just synchronic (*oaf - oafs*). Only alternating paradigms are subject to the HEA, and an older paradigm that is said to belong to an older language stratum (*loaf - loafs [-vz]*) can be constructed. The statement about which of the paradigms is the "older" one can be directly derived from the synchronic linguistic competence.

3.1.5 Examples

3.1.5.1 Ancient Greek ἔχω - ἕξω - ἔσχον

For a better understanding, the morphophonemic method as the prime example of IR is illustrated here through two concrete examples. The first example is particularly popular in the literature (Bonfante 1945:84–85, Pisani 1938:33) and reconstructs the ancient Greek sound change *s > h* using the example of *ἔχω* *$\acute{e}k^h\bar{o}$* 'to have'. In the first step, all wordforms are collected and broken down into their morphemes (Tab. 3.A).

[16] Ringe (2003:254)

Table 3.A. Partial paradigm of the ancient Greek word ἔχω *ékʰō* 'to have' and the morphological analysis of relevant wordforms.

Word form	Grammar	Morphological Analysis
ἔχω /ékʰɔː/	1st sg. pres. ind. act.	*ékʰ* [root] = ɔː [1st sg. pres.]
ἕξω /héksɔː/	1st sg. fut. ind. act.	*hék* [root] = s [fut.] =ɔː [1st sg. pres.]
εἶχον /êːkʰon/	1st/3rd sg. ind. impf. act.	e [augment] = *i-kʰ* [root] = on [1st sg. prt.]
ἔσχον /éskʰon/	1st/3rd sg. ind. aor. act.	é [augment] = *skʰ* [root] = on [1st sg. prt.]
ἔσχηκα /éskʰɛːka/	1st sg. ind. perf. act.	é [augment] = *skʰɛː* [root] = k [perf.] = a[1st sg. perf.]
…	…	…

The HEA assumes that all allomorphs of εχ ekh originate from a "pre-morpheme." To determine the phonetic form of this morpheme, it is worthwhile for the time being to consider all allomorphs as the original form and apply "regular" phonological rules to these wordforms (exemplified for the first sg. ind. act. present, future, and aorist in Tab. 3.B).

The application of the synchronic phonological rules already explains the alternation of *kʰ* ~ *k*. Furthermore, the lack of aspiration in the future forms is caused by the loss of aspiration before *s*. Only the development *kʰ* > *k* can be assumed here; the other direction — an aspiration before a vowel — cannot be inferred by any synchronic rule. By assuming a pre-form **hekʰ*, the aspiration-dissimilation as a second phonological rule is revealed: Each aspirated sound is deaspirated if the following syllable begins with an aspirated sound (Rix 1992:107).

Table 3.B. Assuming four different roots (*ekh*, *hek*, *skh*, and *hekh*). one would expect these wordforms for the first sg. ind. act. present, future, and aorist of the word ἔχω *ékhō* ‘to have’ (Tab. 3.A). Wordforms with † are wordforms, one would expect after applying phonological rules.

Root	Present	Future	Aorist
ek^h	ek^h = ɔ:	*ek^h = s = ɔ: > †eksɔ:	*é = ek^h = on > †ɛ:k^hon
hek	*hek = ɔ: > †ekɔ:	hek = s = ɔ:	*é = hek = on > †hɛ:kon
sk^h	*sk^h = ɔ:	*sk^h = s = ɔ: > †sksɔ:	é = sk^h = on
*hek^h	*hek^h = ɔ: > ek^hɔ:	*hek^h = s = ɔ: > heksɔ:	*é = hek^h = on > †hɛ:kon

The alternation of the allomorphs *hek^h ~ sk^h remains unexplained and cannot be derived by synchronic phonological rules. In the next step, the HEA suggests *h(e)k^h or *s(e)k^h as possible pre-morphemes. The decision between both morphemes is made by an inferring method. Bonfante (1945:85) argues that the direction *s* > *h* is the “usual phonemic change,” while *h* > *s* is unlikely. Accordingly, *sek^h or sk^h is to be assumed as the pre-morpheme. The root structure indicates a sound change conditioned by the subsequent sound: *hV* and *sC*. To explain the absence and presence respectively of *e*, a second sound change (either a syncope or an epenthesis of *e*) is required to have taken place before *s* > *h*. Therefore, the condition of this sound change cannot be determined internally.[17]

Accordingly, we have reconstructed the wordform *s(e)k^h, which is attributed to “Pre-Greek.” Furthermore, we were able to establish what is referred to as a *relative chronology* (i.e., to determine different phases of Pre-Greek). The alternation k^h ~ *k* belongs to the most recent layer of change, preceded by the change s > h, which is historically more recent than the vowel alternation.

3.1.5.2 Latin nix - nivis - ninguit

A second example of IR can be found in some standard works of Indo-European studies (Beekes 1990:131, Meier-Brügger 2010:173) and illustrates the process with the Latin word for snow, *nix* (= *nik-s,* nom. sg.) ~ *nivis* (gen. sg.). The noun is compared with the verbal nasal-infix present *ni-n-gu-it* ‘it snows’. According to the HEA, all three stems *nigu-*, *nik-* and *niv-* are based on one common root. As a preliminary form, we assume *niX with an unknown sound X, which has developed to

[17] The reason for this alternation is the Indo-European ablaut (zero-grade).

k, *v* and *gu* in the given phonetic environments. Identifying *X* as one of these sounds and applying Latin phonological rules, we would assume the wordforms in Tab. 3.C. The equation of *X* with *gu* seems to be the most plausible here but requires an additional sound change *gu* > *v*. Thus, *nig^w- can be determined to be the pre-form.

Table 3.C. Assuming three different consonants (*k*, *v*, and *gu*) as last root consonant, one would expect these wordforms for nom. sg. and gen. sg. of the Latin noun *nix* 'snow' and the verb *ninguere* 'to snow'.

	Nom. Sg.	Gen. Sg.	Verb
k	*nik-s*	nik-is	nink-it
v	niv-s > nip-s	*niv-is*	ninv-it
gu	nigu-s > *nik-s*	nigu-is	*ningu-it*

In this example, an inferring method that selects the most plausible sound change with regard to phonetic context is used. A complication with this method is that often, none of the "implausible" sound changes are phonetically impossible. Labiovelar sounds such as *v* could have changed to g^w after nasal consonants (which is referred to as *sharpening*). Ultimately, the empirically more frequent sound change g^w > *v* and the morphophonemic change g^w > *k* / _*s* yield the correct interpretation. The fact that the form *nig^w- is neither the correct Proto-Italic nor the Proto-Indo-European root can only be resolved through comparative reconstruction. The Comparative Method shows that the correct PIE sound would have been the sound *g^{wh}, but its aspirated feature is not reflected in Latin; thus, it cannot be reconstructed internally. This fact makes it clear that IRs are to be understood more approximatively than reconstructions made by the Comparative Method.

3.2 Formal Structure of Internal Reconstruction

3.2.1 Internal Reconstruction as a Reconstruction Method

The formal structure of each IR follows the example of comparative reconstruction, which has already been presented in Chapter 2. First, the basic hypothesis will be discarded or ascertained in the further course of the sound-change reconstruction. The BASIC HYPOTHESIS equates two linguistic units (usually phonemes or morphemes) on a diachronic level, provided that they meet the conditions of the hypothesis. The reliability of the falsification of the basic hypothesis therefore reflects, to a certain extent, the reliability of the method itself. It is no coincidence that many critics of IR begin their criticism with this point although the method is similar to the Comparative Method. Since the basic hypothesis is defined only for linguistic units within a language, the words in question are referred to as "internal cognates."

From the basic hypothesis, a theory (e.g., *t* corresponds to *d*) that can be falsified via an induction procedure (see Sect. 2.1) is derived. In practice, loanwords, analogy, and other factors may lead to erroneous falsification, so the basic hypothesis is implicitly extended by AUXILIARY PREMISES. Linguists integrate these premises by editing the cognate pairs (e.g., by excluding all words that could be loanwords in accordance with the premise). However, the auxiliary premise can also distort the results since the decision about loanwords is left up to the linguist, who could "explain away" counterexamples. As a rule, loanwords are inferred by comparison with other languages, an approach that is excluded in IR. The empirical data sets of an IR are often too small for an inductive reasoning process, and analogy can result in the hasty falsification of a sound change. As a result, literature on IR does not usually perform a detailed process of inductive reasoning; one typically settles for a deductive application of IR to individual cases. This procedure ensures that the basic hypothesis does not lose its hypothetical nature. Instead, self-formulated auxiliary premises serve to falsify the basis hypothesis.

In this thesis, the procedure of IR, just as that of the Comparative Method, is understood as an inductive reasoning method supported by additional abductive hypotheses (basic hypothesis, inferring methods). Hajnal (2016) had a contrary view, regarding IR primarily as an abductive method (2016:439), while seeing the Comparative Method as an inductive method with abductive premises (2016:442). The abductive nature of IR is due to the fact that it can only determine the original sound of an alternation with the most plausible assumptions (cf. Hajnal 2016:439). However, this view overlooks that even the Comparative Method can determine the original sound of a sound correspondence using only abductive inferring methods. In this respect, the procedure of IR is identical to that of the Comparative Method; they merely work with different data and basic hypotheses.

3.2.2 Is the Basic Hypothesis Solid?

Comparative and External Reconstruction start from the basic hypothesis of a genetic kinship or loan influence of the words to be compared. These two hypotheses may compete for a word pair *a-b* of languages *A* and *B*, but there are often arguments for or against one of the hypotheses. In addition, there may be no relationship at all between both words, and their phonetic similarity may be due to chance (e.g., Latin *deus* and Greek θεός *theós*).

IR's basic hypothesis — as well as that of External Reconstruction — competes with alternative hypotheses in determining internal cognates. Just as a language can take a word externally from another language, a paradigm can take a wordform internally from another paradigm (i.e., *assumption of suppletion*).[18] On the other hand, it is uncertain whether two allomorphs also randomly correspond to each other (*assumption of pure coincidence*, i.e., null hypothesis). In these cases, an allomorph would have been completely re-formed; that is, it was derived neither from an allomorph (hypothesis 1: HEA) nor from another lexeme (hypothesis 2: assumption of suppletion). Such cases are probably only conceivable outside of morphophonemic IR (see Sect. 4.2).

Table 3.D. Possible hypotheses for the similarity of two words of different languages (Comparative Method) and the same language (Morphophonemic IR).

Comparative Method	Morphophonemic IR
Hypothesis of genetic kinship	Hypothesis of etymological allomorphy
Hypothesis of loan influence	Hypothesis of suppletion
Assumption of pure coincidence	-

In addition to loan influence and genetic relationship, other factors may cause phonetic similarity, including babble words and onomatopoeias. In the view of IR, such factors are to be subsumed under the term "suppletion." In paradigms, however, babbly words and onomatopoeias usually occur less but not infrequently within a semantic field. An example would be the German words *kikeriki* 'cock-a-doodle-do' and *Küken* 'chick', both of which are descended from onomatopoeia.[19]

The hypothesis of genetic kinship can be rejected for cognates if the pair violates the deductive conclusions of inductive reasoning. From the inductive reasoning procedure, the sound correspondence between the PIE **ē* (< **eh_1* and **ē*) and the Gothic *ē* results. This correspondence, deductively applied to the word PIE **h_3rḗǵ-s* 'king' results in the Gothic wordform †*rēk-s*. The attested wordform *reiks* /rīk-s/ contradicts the comparative hypothesis and argues for the assumption of a loanword from

[18] Anttila (1973:329), on the other hand, regarded analogy as "borrowing from within." This comparison is only true on phonetic level, but the ways they work are quite different.
[19] The German word *Küken* goes back to the onomatopoeic root **kuk* (Kluge 2011:547).

Celtic, where the sound change PIE *$\bar{e}$ > Proto-Celtic *$\bar{\imath}$ is proven. If the borrowing had been carried out before the Celtic sound change, the phylogenetic hypothesis could not have been rejected since the phonetic form would have coincided with the expected phonetic form. The hypothesis of borrowing, just like the assumption of coincidence, can therefore never be completely rejected.

Comparable cases are conceivable for IR, especially when using the derivational method (see Sect. 4.2.2). As with the Comparative Method, alternative hypotheses of IR can never be completely discarded. Nonetheless, the alternative hypotheses of both reconstruction methods only take hold when the basic hypothesis is rejected. If a suppletive wordform coincides more or less by chance with the "regular" derivation (e.g., because it was reanalyzed as derivation of the word), the suppletion cannot be recognized as such (e.g., German *gehen* 'to go (Pres.)' and *ging* 'went (prt.)', cf. DUDEN 307). Similarly, dialectal and diachronic borrowings, such as Italian *pieve* 'a rural parish church' and *plebe* 'common people' (both from Latin *plēb-s*) can only be identified in few cases (see Hoenigswald 1965:68).

A fundamental problem inherent to IR is that the inductive process can only take effect when it is supported by sufficient data. For synchronic phonological rules, this condition is certainly met, as they occur in all affected wordforms. Older sound changes may have been obscured by analogy or more recent sound changes, shrinking the available data to a minimum; here, the risk of circular reasoning may exist.

In the literature, the question of hypothesis rejection has rarely been raised. Various proposals have been put forward to circumvent the problem of insufficient data. Marchand (1956:247) uses a formal constraint for this purpose: Allomorphs are said to be suppletive wordforms if they "have the same phonemes in as many positions as they have different ones, and these phonemes must be in the same order." Other scholars, such as Hock (2013:9), dispense with a "hard" constraint and assume that the fewer sound correspondences there are, the less likely the basic hypothesis holds true.

3.2.3 Differences with Other Linguistic Methods

Table 3.E. Differences between IR, Comparative Method (CM), Morphophonemic Analysis and Phoneme Analysis according to Anttila (1968:169).

<table>
<tr><th colspan="3">Input</th><th colspan="3">Output</th></tr>
<tr><td>number of languages</td><td>phonetic units</td><td>conditioning</td><td>name of procedure</td><td>name of units</td><td>history vs. description</td></tr>
<tr><td rowspan="2">single a</td><td rowspan="3">sets of corresponding sound units e</td><td rowspan="2">grammatical f</td><td>IR</td><td>pre-phoneme</td><td>H</td></tr>
<tr><td>Morphophonemic analysis</td><td>morphophoneme</td><td>D</td></tr>
<tr><td>many d</td><td rowspan="2">phonetic c</td><td>CM</td><td>proto-phoneme</td><td>H</td></tr>
<tr><td>single a</td><td>phones b</td><td>Phonemic analysis</td><td>phoneme</td><td>D</td></tr>
</table>

Not only shows IR similarities to the Comparative Method but also to the morphophonemic and phoneme analysis. Anttila (1968:169) observed the differences between these four methods in the differences in their “input” data and their target or purpose (“output”), as illustrated in Tab. 3.E. In this account, the similarity between IR and morphophonemic analysis is salient, differing only in purpose (*history vs. description*; see also Sect. 6.1.1). A strict separation between the four methods can only be found in the final result (*name of units*). The results can, in turn, serve as inputs for other methods:

> *[T]he output units can be fed again into the machine, e.g. by putting the phonemes back into boxes e and b. […] Thus one usually applies the comparative method (4.) to phonemes […] Internal reconstruction […] can also be applied to a reconstructed protolanguage […] If we apply the comparative method (4.) on the morphophonemes (result of 1.) of the two*

> *languages, the result (3.) is more similar to internal reconstruction in that no gradation results […] this would be the result of taking to the ‚inherent order of application‘ of internal reconstruction first*[20]

3.3 Motivation for Internal Reconstruction

The Comparative Method was established as the basic procedure of reconstructing historical linguistics, beginning with the findings of the neogrammarians. The possibility of falsification provides a positivistic feature to historical linguistics and enables objective and reproducible research.[21] IR could not occupy a comparable position. While at the beginning, IR seemed to indicate the possibility of reconstructing older language stages (or even the oldest language stages, cf. Hockett 1958:462), in today’s introductions, emphasizing the limitations of IR occurs on a regular basis (e.g., see Crowley and Bowern 2010 Chapter 7.3, Campbell 2013 Chapter 8.4). In introductions to Indo-European studies, the topic of IR is hardly addressed (e.g., Mallory and Adams 2006, Baldi 1983). Unlike the neogrammarian hypothesis of the “exceptionlessness of sound laws,” the HEA has never been developed into a linguistic theorem. The attempt to master its presupposition — to recognize and identify suppletion synchronically — eventually ended in resignation. The scope of IR’s application was, therefore, pushed into the periphery, where the Comparative Method reaches its limits.

3.3.1 Isolated Languages

One of the tasks of IR is its application to languages with no known linguistic relationship and no linguistic-historical documentation (cf. Chafe 1959:478, Birnbaum 1970:98), which is sometimes seen as “its most useful” benefit (Austerlitz 1986:183). However, what linguistic significance IR actually has in this application remains an open question. In fact, especially for languages with much loan influence, such as Basque, External Reconstruction seems to be a far more important and reliable basis for reconstructing sound laws. Thus, it is not by chance that loanwords

[20] Anttila (1968:169–170)
[21] Nevertheless, the Comparative Method is not able to verify a theory (i.e., a postulated sound change). As Popper (1959) pointed out, this is generally not possible in science from an epistemological view.

usually serve as examples of Basque sound changes in standard works (e.g., Michelena 1977:251 or 257, among others). In addition to the greater reliability of External Reconstruction, one reason for loanwords being examples of Basque sound changes may be the agglutinative character of Basque. However, approaches that focus on IR with Basque can also be found, especially in more recent works (Blevins 2018, Martínez-Areta 2013:321). IR seems far more popular for languages with little to no written tradition. The method is therefore frequently applied for indigenous languages in the Americas and Africa (cf. Sect. 5.9).

In those papers, IR often also serves as a means of connecting an isolated language to another language or of detecting macro-families (e.g., *Hokan languages* in Haas 1980 and Langdon 1997). Here, the "synchronic arrangement of morphemes" that linguists try to use IR to overcome can make a connection difficult (Haas 1980:67). Especially in Korean-Japanese studies, IR has therefore been repeatedly used (Unger 2000, Martin 1996). The detection of identical sound changes in the (proto-)languages concerned provides an indication of potential kinship because "one and the same sound change can be supported by both K[orean]-J[apanese] matches and a CONSISTENT internal reconstruction of verb paradigms is exactly the kind of intersection of methods one hopes to find generally in historical linguistics. [original emphasis, A.B.]" (Unger 2000:679).

A special application of IR comes from the work of Ivano and Toporov (1961:275–276 and 303–304), who compared internally reconstructed relative chronologies of sound changes to draw conclusions about genetic relationships between Slavic and Baltic languages. They showed that internally reconstructed Pre-Slavic can be derived from Pre-Baltic but not the other way around.

3.3.2 Reconstruction of More Recent Language Stages

IR is particularly suitable for the reconstruction of more recent language history (cf. Chafe 1959:478, Birnbaum 1970:98). The Comparative Method does not allow to reconstruct more than one stage (the proto-language) and is consequently unsuitable for this application (cf. Shevelov 1964:6–9). Since IR tends to establish relative chronology of sound changes rather than wordform reconstructions, this is sometimes regarded as the "prime task" (Anttila 1973:325) of IR. The Comparative Method's lack of determining relative chronology is often used as an argument for IR (cf. Anttila 1968:159).

However, in this way, IR also competes with the External Method, which also provides an insight into the time between the proto-language and the beginning of the written tradition. An example of this is Euler and Badenheuer (2009), who does not use the term "internal reconstruction" in their book about Pre-Germanic at all, while loanwords are very well invoked, likely because of the temporal indeterminacy of the IR's pre-forms. In the simplest case, IR provides no more than the time dimension "pre-linguistic," while External Reconstruction postulates a more concrete time period.

3.3.3 Reconstruction of Pre-Proto-Languages

One of the most popular applications of IR is the reconstruction of pre-proto-language stages (cf. Birnbaum 1970:98–99, Fox 1995:211). In fact, IR is often the only means of reconstruction for this application, as other reconstruction methods have limits due to their time depths. However, it is precisely this point that has led to the criticism of "glottogony" in IR (see Sect. 6.3). The possibility of gaining new time depths with IR was questioned by Prosdocimi. He (1977:95) assumed that the application of IR to already reconstructed proto-languages does not create new time depth because "this kind of reconstruction presupposes comparison." However, he does not further articulate his argument.

Instead, the *incompleteness of proto-languages* and the *subjective linguist's influences* seem to constitute an obstacle to this application of IR. To apply IR "for a reliable ultimate answer one needs the total language, the total grammar, as input" (Anttila 1973:318), which is why some linguists have rejected the application of IR to proto-languages (e.g., Hoenigswald 1974:188). Anttila (1973:336) even saw "the relation of this machinery to the material culture" as necessary to gain a complete image of a language. If a language is not fully present, alternations may be erroneously assumed because the corrective non-alternating pairs may not be randomly attested or reconstructed. A more practical problem is the subjective influences of reconstructed proto-languages. Most handbooks or dictionaries of proto-languages are more or less implicitly based on phonological or morphological assumptions. This already includes the assumption of missing open roots or a phoneme /a/ in Proto-Indo-European. An IR based on these data would only be able to reproduce those assumptions.

In addition, reconstructions based on proto-languages have no further means of verifying the results. The advantage of IR is that it can be applied even at very high time depths in contrast to other methods; at the same time, however, this also constitutes a disadvantage. Lass (1975:18) therefore came to the assessment that "the place where IR is of crucial importance is where the evidence that COULD test it is irretrievably lost [original emphasis, A.B.].” Through such a drastic assessment, there is a threat of reinterpreting IR as a speculative method without adding scientific value.

3.3.4 Unclassified Languages

Isolated, extinct languages for which no bilingual inscriptions or comparable material is accessible are largely incomprehensible today. Etruscan and Minoan are among the best-known languages of this kind. In works on these languages, the recognition of related morphemes plays a major role. The question of whether the wordforms *xyz* and *xya* belong to one-morpheme *xy* or to two different morphemes can contribute decisively to the understanding of the language. IR can be helpful in determining alternations. In practice, the problem with this approach lies in the fact that the morphophonemic method of IR ideally presupposes the identification of allomorphs. This basic condition may not hold for many unintelligible languages.

3.3.5 Review of the Results of Other Reconstruction Methods

In addition to the usual procedure of controlling the results of IR by using the corresponding results of the Comparative Method, IR is conversely used to verify the results of the Comparative Method (cf. Hoenigswald 1944:79 and 1950:362). In fact, Marchand (1956:245) criticized the fact that the use of IR primarily serves to confirm the results of other reconstruction methods. This is especially true in cases where possible reconstructions that are incorrect must be ruled out. If two sister languages A and B show different reflexes of the phoneme /x/, the proto-sound can be determined by IR (see Fig. 3.1). Especially when comparing only a few languages, IR can often only indicate the correct original sound (cf. Hoenigswald 1950:363). In language families with many languages, a feature only attested in one language can be detected in other languages by means of IR and thus — if necessary — correct a proto-language reconstruction (cf. Shevelov 1964:5). On a lexical level (e.g., in determining an inherited animal name), morphologically complex wordforms (such as ‘blackbird’ for ‘merle’) or taboo words can be identified as innovations, while languages with non-analyzable terms likely preserved inherited words. However, loanwords may render this rule useless.

3.4 Application with the Comparative Method

In practice, IR is applied in connection with the Comparative Method more often than as the sole reconstruction method. In the second half of the 20th century, a debate arose on the question of whether IR should be applied before or after a comparative reconstruction (cf. Anttila 1968:165–166). In this context, the PRE-COMPARATIVE APPROACH seems to be the older and more widespread view (e.g., Hermann 1907, Chafe 1959:494–495, King 1971:201, Kiparsky 2014:65; cf. the "revised internal reconstruction" according to Unger 2000:660). The purpose of a pre-comparative approach would be to purify the data for the comparative approach:

> *Before comparative analysis is undertaken, each body of synchronic data should be examined carefully to see whether there are not internal clues, mainly morphophonemic, to the structure of the language at a slightly earlier stage. When there are, any possible internal reconstruction should be undertaken, since it will dispose of later innovations in each language, getting them out of the way for a clearer view of the deeper time-perspective of external comparison.*[22]

The alternative view is the POST-COMPARATIVE APPROACH. This is present, for example, in what Hall (1950:9) called "intermediate reconstruction," which compares some languages by means of the Comparative Method and subsequently modifies the result through an "internal reanalysis" and through "evidence of other dialects." The IR here has a corrective function for the Comparative Method (as with Lehmann 1961:24, cf. also the "integrated comparative method" according to Shevelov 1964:6–7 with the purpose of integrating a relative chronology into the Comparative Method).

[22] Hockett (1948:124)

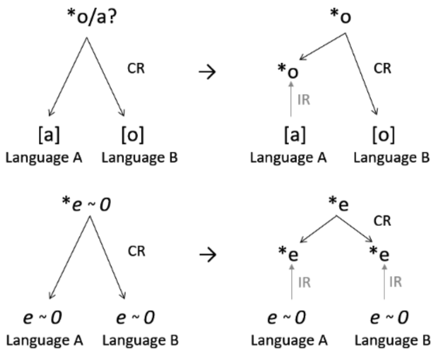

Figure 3.1: Application of IR before a comparative reconstruction (*pre-comparative approach*; upper figure) and after a comparative reconstruction (*post-comparative approach*; lower figure).

Both approaches, however, show potential sources of error. If the pre-comparative approach is used, alternations that occur in all daughter languages and are likely to be of proto-lingual origin can be eliminated before the Comparative Method is applied to the pre-languages. Campbell and Grondona (2007:24) cited the example of a vowel loss attested in both Chulupí and its sister language Maká, which already existed in their proto-language Proto-Matacoan. Since the alternation was already removed in Pre-Chulupí and Pre-Maká, it will also be missing in the reconstructed proto-language. On the other side, erroneous reconstructions may also occur through the post-comparative approach:

> *In some instances, the prior application of internal reconstruction feeds comparative reconstruction, as is the case here with the Delateration sound change (kl > k / word-finally and before certain consonants), which has the result that no Chulupí k which alternates with kl needs to be compared directly with l in the other Matacoan langues; rather the l of these other Matacoan languages can be compared to reconstructed *kl in Pre-Culupí if internal reconstruction is applied before the comparative method.*[23]

[23] Campbell and Grondona (2007:24)

Expert discussions about the question of the correct order finally led to a compromise, which can be attributed to Anttila (1989:274). This compromise called for a test procedure in which a linguist must check the results in both orders of application (cf. Campbell and Grondona 2007:25).

The problem that arises when combining IR with the Comparative Method is ultimately due to the different dimensions of the methods. If the pre-lingual sound change took place before the proto-lingual time, IR must be applied after the Comparative Method; otherwise, the comparative result must be corrected using IR.

3.5 Relative Chronology

The different areas of application lead to the next question of what exactly IR is trying to reconstruct. The internal reconstructability of past sound change depends on various factors. In addition to loanwords, analogies, and neologisms, this also includes the possibility of "overlapping" with more recent sound changes, resulting in a language that synchronically shows relicts of old phonetic alternations (e.g., the PIE *ablaut*) but may not have preserved any residue of more recent sound changes. With regard to its temporal dimension, the method of IR provides only an anachronistic framework (see Sect. 6.3.1), which only allows the designation of the reconstructed language with the auxiliary term "pre-lingual." However, the method theoretically only enables the reconstruction of "non-overlapped" sound changes, which narrows the time depth to a "primarily recent" pre-lingual stage.

The overlapping of older sound changes by younger changes leaves traces in the phonotactics, which may at least partially cancel out an anachronism: The covered sound change can be classified as older *relative* to the other sound change. Thus, a so-called ***relative chronology*** can be postulated between both sound changes, the prerequisite of which is an interference between both sound changes (cf. Chen 1976:251). A relative chronology is often also established for wordforms. Alternating words are assigned to the older stratum, while non-alternating (or no longer alternating) words are attributed to the younger stratum.

The Comparative Method often cannot make a statement about relative chronology with its two-stage model of proto-language and attested stage. Nevertheless, it is possible to create relative chronologies through Comparative and External Reconstruction (see Bremer 1894:15–16, Hoenigswald 1992:31). An absolute chronology is only possible with the help of extra-linguistic knowledge. In the second half of the 20th century, a prevailing view indicated that relative chronology was the main task of IR: "[this method] has its chief goal the recovery of the order or, rather, the relative chronology a set of preceding linguistic changes" (Birnbaum 1970:98). Indeed, IR frequently makes more statements about relative chronologies than about reconstructed wordforms and, conversely, "statements about units can be converted into statements about" relative chronologies (Anttila 1973:325).

In the literature on reconstruction methods for relative chronology, a variant use of the term "internal method" occurs. For example, the (comparatively) reconstructed "protoforms and their modern reflexes" (Chen 1976:209) can already be taken as a given, such that from Lat. *ornus* < **osinus* 'ash tree' a relative chronology of (1) *intervocalic rhoticism* (*s* > *r*) and (2) *syncope* can be assumed. This kind of "internal" reconstruction was documented in the 19th century (e.g., Stolz 1894) but does not allow for any definitional demarcation of the reconstruction methods due to its implicit use of other methods.

3.5.1 Automatic and Non-Automatic Alternations

The anachronistic status of IR's results does not primarily allow any differentiation between synchronic *alternations* and diachronic *sound change*. However, IR's claim to reconstruct older language stages makes it necessary to recognize and classify phonological alternations. The question of the temporal dimension of a recognized sound alternation was therefore an early task of linguists who tried to determine sound changes with the help of IR. Nevertheless, the need for such differentiation has also been called into question. Chafe (1959:481) only considered it necessary for the "reconstruction of a relative chronology of the changes." He argued that productive sound alternances can be regarded as "the results of the most recent sound changes" (Chafe 1959:481). It should be added to this view, however, that synchronic sound changes are not necessarily phonemic changes and that such "sound changes" can be "undone" through simple reanalysis; for example, the German neologism *hamstern* [ˈhamstɐn] 'to hoard' is derived from *Hamster* [ˈhamstɐ] 'hamster' and therefore has never been pronounced with final *-r*. For the first singular form *ich hamster-e* [ˈhamtəʁə], a diachronic view would assume that the historical sound change **əʁ* > *ɐ* \ _[-vowel] was "undone" here, or a new sound change should be assumed to explain the development *ɐ* > *əʁ*.

Instead of *phonological alternations* and *sound changes*, the literature of IR usually uses the terms "automatic" and "non-automatic" (Hoenigswald 1944:84, Chafe 1959:481), "compulsory" and "non-compulsory" (Hoenigswald 1974:192), or (more rarely) "productive" and "unproductive" (Anttila 1973:344). These terms are mostly used interchangeably (e.g., Fox 1995:157) but are sometimes defined differently. Bloomfield (1933:211) coined the term *automatic* as a sub-term for "regular." *Regular* refers to an alternation in which the alternants are distributed according to linguistically recognizable features (*characteristic*) of the environment (e.g., German [x] and [ç]). If the decisive feature is phonetic, Bloomfield speaks of an "automatic" alternation (e.g., final devoicing). An alternation can be "non-automatic" but "regular" if a morphological condition occurs.

Additionally, the terms "automatic" and "compulsory" have also been defined differently. Chafe (1959:482) mentioned two possible definitions of "automaticity." In the first case, an alternation is "automatic" as long as it "can be generalized in terms that apply AT MORPHEME BOUNDARIES to all BASIC ALTERNATES [i.e., allomorphs or morphophonemes; original emphasis, A.B.] containing a certain feature." Accordingly, the ancient Greek rule s > 0 / V_V, which gave rise to alternations of *εἶ eī* (< **es-i*) 'you (sg.) are' ~ *ἐστί es-tí* 'it is', continued to be "automatic" after intervocalic sigmas arose from **ti*. These secondary sigmas can only be found at morpheme boundaries, as in the word *am-bros-ía* 'immortality'; within the morpheme itself, *s* is not intervocalic. In the second case, a rule must apply to the entire phonotactics of a language in order to be called "automatic." Chafe (1959:482) preferred the latter definition as it seemed "more useful" in application. To distinguish between both of Chafe's definitions, Hoenigswald (1974:192) eschewed the term "automatic" in his later works and defined the first cases as "regular" and the second ones as "compulsory" alternations. On the other hand, Kurylowicz (1964:14) excluded allophones from "automatic" alternations and restricted his definition to cases in which different phonemes alternate with each other.

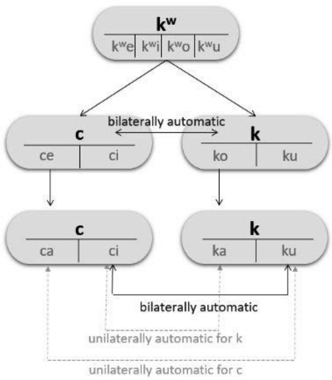

Figure 3.2: Illustration of the automaticity of *k* and *c* in Vedic Sanskrit (last stage). The Proto-Indo-European sound $*k^w$ became *c* and *k* in Pre-Sanskrit, which were bilaterally automatic allophones.

In the literature on IR, different types of automaticity have been distinguished. If a sound *s* occurs in environment *X* in which the sound *t* does not occur, this position is called BILATERALLY AUTOMATIC. An alternation "involving *m*-in-1 is UNILATERALLY AUTOMATIC with the *m*-allomorph as a BASE, in the sense that the *n*-allomorph occurs in its stead where, and only where, the phonemic system precludes the *m*-allomorph [original emphasis, A.B.]" (Hoenigswald 1965:101). This differentiation can be illustrated by the Sanskrit palatal and velar plosives. From the PIE sound sequences $*k^we$ $*k^wi$ $*k^wu$ $*k^wo$, the Pre-Vedic Sanskrit sequences **ce* **ci* **ku* **ko* arose through palatalization and delabialization. Both *c* and *k* occurred in different environments and are *bilaterally automatic*. Due to the subsequent merge of *e* and *o* in Sanskrit, the sound sequences *ca ci ku ka* occurred. A new alternation combination took form: Before *i* und *u*, *c* and *k* continued to be *bilaterally automatic*. Before *i* and *a*, they were unilaterally automatic with base *k* since *k* still did not occur before *i*, but *c* did occur before *a*. Conversely, before *u* and *a* there was a unilaterally automatic alternation with base *c* (*ca* and *ku* but no **cu*; cp. Fig. 3.2). Because of the presence of both sounds before *a*, the alternation of *c* and *k* is generally considered to be "NON-AUTOMATIC [...] and IRREGULAR in the sense that the morphs containing the feature [...] Skt. -*a* [...] must be either listed or suitably identified by a morphophonemic notation [original emphasis, A.B.]" (Hoenigswald 1965:103; e.g., a_1 and a_2).

3.5.2 Establishing Relative Chronologies

To establish a relative chronology, several proposals have been put forward in the literature. The concept of *automaticity* played an important role in the development of chronologies, especially among structuralist linguists, such as Chafe. A relative chronology based on automaticity is only possible "when changes interfere with each other in that the outcome of one is either the input or the environment of another" (Anttila 1973:330). Other concepts are rarely used to establishing relative chronologies. A selection of such methods is presented in this section.

Methods for a relative chronology of wordforms were stated by Givón (1999a:121–122), such as the rule "[t]he smaller a morpheme is, the older it is." However, these were actually methods of a diachronic projection (see Sect. 2.3.2), which do not truly constitute a reconstruction.

3.5.2.1 Relative Chronology by Automaticity

The concept of automaticity enables a distinction between diachronic and synchronic alternations and is therefore the first and easiest way to establish a chronology between sound changes (cf. Chafe 1959). The productive synchronic sound changes belong to the temporally younger stratum, while the non-productive ones belong to the temporally older stratum. However, this rule only seems valid if the "age" of a sound change is defined as its time of transition to unproductivity. Such a distinction plays a role for sound changes that remain productive over several generations. An example is the German final devoicing, which already existed in Middle High German and has remained productive until today (cf. also Chen 1976:253). After its first occurrence, further sound changes that are now unproductive intervened, but such changes should be classified as older according to Chafe's rule.

Cases for which the unproductive alternation became productive in later times are rare but possible in principle. For example, Mehendale (1963:44) cited the Vedic Sanskrit alternation of *ś* and *ṭ*, which is automatically before consonants only for individual paradigms but not for the whole language: the alternation is shown by the paradigm of *viś* 'settlement' (instr. sg.: *viś-ā́* : nom. sg.: *víṭ* : dat. pl. *viḍ-bhyás*[24]) but is missing in compounds such as *viś-páti* 'lord of the house'. Consequently, the rule is considered non-automatic in Vedic Sanskrit. In later stages of Sanskrit, this alternation was extended to the whole language: *viś-páti* became *viṭ-pati*.

From the coexistence of alternating forms and non-alternating forms (referred to as *invariances*), an unproductivity of an alternation is usually inferred since it has been eliminated, for example, through analogies (e.g., *leaf-leaves, knife-knives* vs. *roof-roofs*.). Anttila (1973:336) saw no indication for a relative chronology in such a coexistence alone since an alternation is productive at the beginning as an innovation and, as such, younger than the old invariant forms. He (1973:336, cf. also 1973:345) therefore called for two additional conditions:

- when the alternation "is older than invariance when it correlates with basic core vocabulary" and
- "one needs the total grammatical machinery and the relation to the material culture for a reliable answer."

However, this limitation seems to play a more important role at the morphological level than at the phonological level. If a morphophonemic alternation occurs (e.g.,

---

[24] The sound *ś* before *bh* normally becomes voiced *ḍ*.

final devoicing), it applies to all words and automatically eliminates invariances. Derivational alternations, however, can be borrowed (e.g., *co**n**-stant* and *co**m**-pley* or *co**m**-puter*). In those cases, a comparison with basic vocabulary can indeed be helpful (cf. ***in**-stead* and ***in**-put* without such an *n-m*-Alternation). Chafe (1959:481) argued that more than two stages (synchronic-diachronic) are possible for a relative chronology in means of automaticity if the process is applied "recursively":

> *[T]he phoneme correspondences extracted from automatic alternations can be used directly as the basis for reconstruction [...] Now at this first reconstructed stage it usually happens that certain alternations not automatic in the attested language were still automatic. On the basis of these alternations it is possible to reconstruct forms two stages removed from the attested language, and, barring other complications, the procedure can be continued in this way from each historical stage to alone that preceded it.*[25]

This case is a "reconstruction on the basis of a reconstruction," which incurs the same problems as pre-proto-language reconstructions do. These are based on the question of whether a closed synchronic language system can be achieved through individual reconstructions (for discussion see Eichner 1988:21–25).

Relative chronologies that use automaticity have the disadvantage of often corresponding to the synchronic order of phonological rules (see Sect. 6.1). An example of such correspondence is given in Fig. 3.3 with a Finnish phonological rule order. Similar to these synchronic rule orders, the order of relative chronologies is not infrequently ambiguous and "rather random," as we can capture it "only in a few favorable cases" (Anttila 1973:330–331).

1. *vete* —e > i / _#→ *veti*
2. *veti* —t > s / _i#→ *vesi*
3. *veten* —t > d→ *veden*

or 3. before 2.

Figure 3.3: Sample of a relative chronology using the example of the Finnish stem *vete* 'water' and its alternating wordforms *vesi* : *vetØtä* : *veden*. (Anttila 1973:330).

[25] Chafe (1959:481)

3.5.2.2 Relative Chronology by Parallelism

The assumption of parallel developments of sounds within the same sound class may indicate another possibility for establishing a relative chronology (cf. Chafe 1959:486). This hypothesis results from the observation that sound changes tend to be initiated less at the phonetic level than at the sub-phonemic level of distinctive features. Thus, final devoicing is preceded by slowly decreasing voicing at the word's final position, which consequently may naturally affect all voiced consonants. This effect can be used for this method. The PIE palatals *ḱ *ǵ and *ǵ^{h} developed into Skt. *ś*, *j*, and *h*, respectively. For this development, an intermediate step *ǵ >*ź > j and *ǵ^{h} > ź^{h}> h was assumed for the voiced palatals in order to be able to construct a parallel development *ḱ > ś. The procedure is more complex in the case of conditioned sound change. Due to the loss of the triggering factors, the alternations of [f] and [v] in modern English can hardly be determined (e.g., in *knife* : *knives*). However, a parallel similarity to the alternation of [θ] and [ð] can be noted here, as in *mouth* : *mouths*, so the "linguist may thus assume an analogous history for /f/ ~ /v/" (Hoenigswald 1944:85). The difficulty pertains to identifying the common features of the environment. Ultimately, the method can only be applied if the decisive features of the sound change can be determined.

It is unclear whether it is actually possible to identify these features synchronically. Even if the condition has been determined correctly, parallel development does not necessarily have to be present. In particular, sound changes in the articulation series *labial-alveolar-guttural* sometimes show a separated development of the gutturals (e.g., the so-called *inner-German consonant weakening* that made voiceless plosives to voiceless lenis plosives; cf. Mitzka 1967:254). One reason for this special development could be that not the feature *lenis* or *voiced*, but another sub-phonemic feature may have been a trigger of the change (cf. Werner 1972:44).[26]

An alternative explanation does not assume the initiation of a sound change by distinctive features but the beginning of a sound change in one specific environment, which then gradually expands to other environments and sounds (*rule extension,* according to Chen 1976:214, see also Hoenigswald 1992:25). For the abovementioned example of plosive series, we would then assume that weaking began with the labial or alveolar plosive and then gradually expanded without completely affecting the gutturals in all environments. In this case, a parallel chronology cannot be assumed.

[26] Instead of the series *labial-alveolar-guttural*, Werner (1972:44) described this sound change using the series *low diffuse-high diffuse-compact*.

At the morphological or syntactic level, the possibility of a more recent analogical formation is added. The unproductive *ablaut*-preterit in the Germanic languages (e.g., *sing-sang-sung* or *forget-forgot-forgotten*) can be interpreted internally as a formerly productive preterite formation in Pre-Germanic but can also be an analogical formation or coincidence in individual cases. Relative chronologies for word-forms therefore remains uncertain. For syntactic structures, Bauer (2009:28) suggested typological studies as a potential help “because the co-occurrence of similar structures in other languages may point to relatedness.”

3.5.2.3 Relative Chronology by Means of Missing Sound Changes in Other Words
Another popular method for establishing relative chronologies is the lack of sound changes in other words, which suggests antecedent dating (cf. Chafe 1959:486). A reconstructed nasal loss in the final position no longer allows any final nasals in the subsequent stage; in that case, any nasal in that position must have originated in more recent sound changes. This method applies here, just as with the chronology by automaticity, but the two sound changes are neither synchronic nor automatic.

3.5.2.4 Relative Chronology by Two Alternations Within One Paradigm
Another approach has been suggested by Mehendale (1963:44). If two different alternations are found at the same position in one paradigm, a relative chronology between them can be established. This is possible if one of both alternations “concerns phonemes which are phonetically closer, then it may be said that the phonetic change which led to this alternation occurred earlier than the one which led to the alternation between phonemes not so close” (Mehendale 1963:44). An example with the Sanskrit paradigm for *viś* ‘settlement’ is given in Tab. 3.F.

Table 3.F. Partial paradigm of Vedic Skt *viś* ‘settlement’ with alternating *ś* : *ṭ* : *ḍ*.

	Singular	**Plural**
Nominative	*víṭ*	*víś-as*
Accusative	*víś-am*	*víś-as*
Instrumental	*viś-ā́*	*viḍ-bhís*

According to the basic hypothesis, it can be assumed that the *ś, ṭ,* and *ḍ* trace back to the same pre-phoneme. This sound is determined on the basis of phonetic similarity: *ś* is phonetically closer to *ṭ* than to *ḍ*. Therefore, a relative chronology is assumed with *ś* > *ṭ,* then *ś* > *ḍ* or *ṭ* > *ḍ* (cf. Mehendale 1963:44). Mehendale already assumed that the pre-phoneme was *ś* and therefore asked for the most similar sound to it. In practice, this often cannot be easily determined. The assumption of the original phoneme decides the subsequent procedure. Another example is given by the comparative paradigm of the German *hoch* 'high' in Tab. 3.G. If we assume a pre-phoneme *ch* /x/, the relative chronology would be *x* > *k*, then *x* > *h*. If we assume a pre-phoneme *h*, the relative chronology would be *h* > *x*, then *x* > *k*. However, the correct sequence is *x* > *h* (before vowel), then *x* > *k*.

Table 3.G. Comparative paradigm of German hoch 'settlement' with alternating ch : h : k.

Positive	Comparative	Superlative
*ho*ch	*hö*h*er*	*am höchsten /hø*k*stən/*

Another problem with this approach is the lack of a definition for phonetic similarity. If none of the three sounds is identical to the original pre-phoneme, the approach makes no sense. For example, if the *ablaut* in English *sing : sang : sung* is considered with the assumption of a pre-phoneme *i*, what would be the phonetically more similar sound? Both possible chronologies (*i* > *u,* then *i* > *a* or alternatively, *i* > *a,* then *i* > *u*) would be incorrect. A relative chronology for the PIE *ablaut* (*e* : *o* : Ø) can be established in other ways: at least the zero-grade Ø can be traced back to an unproductive sound change (cf. Tichy 2009:14) and therefore seems to be the younger sound change.

3.6 The Concept of Reconstructability of Phonetic Internal Reconstruction

The Comparative Method claims to reconstruct the oldest common language stage (i.e., proto-language) of the compared languages. In contrast, IR no linguistic "target stage" and the reconstructability of a sound change depends decisively on its development. However, a language stage cannot be inferred immediately from this developmental stage. While the English *th*-phonemes still show a phonotactic distribution, which indicates ancient sound change (see Hoenigswald 1944:83), the more recent sound change of the *Great Vowel Shift* can hardly be deduced from Modern English. This circumstance led to a typology of sound changes with regard to their different reconstructability (see Sect. 4.1).

3.6.1 Why is it Possible to Reconstruct Internally?

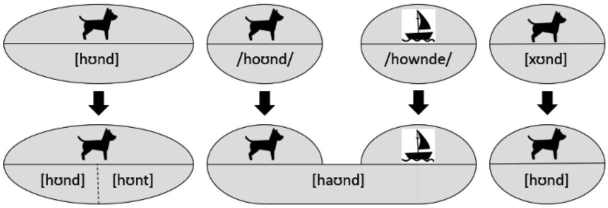

Figure 3.4: Sound changes may have different influences on the *signifié*. On the left, the morpheme for the German word *Hund* 'dog' split through final devoicing. In the middle, the homonymy of English *hound* arose from the Middle English words *hound* 'dog' and *hownde* 'bib' (< Old Norse *húnn*). On the right, the spirantization of Germanic *x in several Germanic languages is an example of a sound change without any influence.

To adequately address the question of reconstructability, it is useful to understand why it is possible to reconstruct internally at all. In general, a sound change directly influences a language's phonotactics. Since sound change is, in turn, a unilateral phenomenon of the *significant*, three possible consequences for the relationship between *significant* and *signifié* result from a sound change. First, a conditioned sound change can cause an allomorphy, whereby several allomorphs refer to the same *signifié*. In the second case, homonymy can result from a sound change. Thus, one string of phones may refer to two different meanings. Finally, there may be no changes in the relationship between *significant* and *signifié* due to a sound change (cp. Fig. 3.4).

Classical IR methods presume the first case and limit their reconstructable sound changes to conditioned sound changes. Second-type cases require methods including semantics (cf. Sect. 4.3). These are much more difficult to grasp than the former type. The last type leaves (almost) no traces in the language's system and cannot be captured internally at all or may only be captured indicatively.

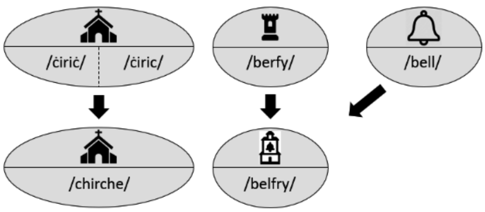

Figure 3.5: Examples of sound replacements. On the left, the paradigmatic levelling of the Anglo-Saxon *ċiriċe* 'church' resulted in the Middle English *chirche*. On the right, the folk-etymological re-interpretation of English *berfry* (< Old French *berfroi* < MHG *bergvrit*) resulted in the wordform *belfry* under the influence of the word *bell*.

Sound replacements must be distinguished from sound changes. In contrast to sound changes, sound replacements are two-sided and therefore directly or indirectly depend on the *signifié*. Indirectly, this happens with analogous replacements within a paradigm, where a reference to an allomorph is created via the meaning that the wordform adapts. In folk etymologies or onomatopoetic phonetic substitutions, the *significant* is *re*analyzed under the influence of another morpheme and *re*-shaped accordingly (cp. Fig. 3.5).

3.6.2 Aims of Internal Reconstruction

According to traditional notions, IR is meant to achieve two goals (Latta 1978:15):

- The description of a part of the language before alternation occurs (i.e., the pre-form, derived in the process of wordform reconstruction)
- The description of the process responsible for the alternation (i.e., the sound change, derived in the process of sound change reconstruction)

In principle, IR reconstructs *pre-forms* (i.e., historical wordforms without determinable time depth and not *proto-forms*). With the lack of time depth, the assignment of a wordform to a concrete language stage is also missing, which would be necessary for reconstructions of proto-languages. Although the knowledge about the non-reconstructability of proto-forms has been known since Hermann (1907:16), a proto-form reconstruction is nevertheless occasionally assumed; both on the side of IR supporters (Hoenigswald 1965:37, Kelly 1960:351–352), as well as on the side of its opponents (Lass 1975:10). However, it could also be concluded from temporal blurring that even preforms cannot be reconstructed. Two sound changes in a wordform inferred from alternations may have taken place at different time, a phenomenon that is dubbed "UNEVEN RESULTS" in the literature (cf. Lounsbury 1953:381 and Hoenigswald 1965:39). Assuming the German umlaut and the sound change *t* > *z* (e.g., via *Kater* ~ *Katze*) can be reconstructed internally, we would postulate the following development for the German alternation *Zwang* ~ *Zwänge* 'compulsion(s)':

- *Zwänge* < **twänge* < **twangi* (first *umlaut*, then *t* > *z*) or
- *Zwänge* < **zwangi* < **twangi* (first *t* > *z*, then *umlaut*)

The pre-form reconstructed in this way has never existed because we could not reconstruct the interposed sound change *Þ* > *d* and *dw* > *tw*. The correct Germanic proto-form was *Þwangi-* (> OHG *(ge)duang* > MHG *twang*). This problem arises when reconstructed sound changes inferred from other data (here *Kater* ~ *Katze*) are deduced to other data and are not IR-specific but a general reconstruction problem already noted by Bremer for the Comparative Method (see Sect. 2.1.2.2). However, IR is generally more susceptible to this and therefore usually avoids a deductive approach in wordform reconstruction. The classical morphophonemic approach of IR would merely reconstruct the umlaut from the alternation: *Zwang* ~ *Zwänge* < Pre-German **zwang-i*.

This pre-form deviates significantly from the historical proto-forms. If one further includes the *ablauting* forms of the verbal paradigms *zwingen* ~ *zwangen* ~ *gezwungen*, an even more pronounced anachronism arises. The same problem may occur with loanwords and analogously levelled wordforms. IR reconstructs a Pre-German **grad* from the German alternation [graːt]:[gradə] 'degree(s)', which has never been in use since the word was borrowed into German only after the beginning of the final devoicing. Conversely, the non-alternance of [vɛrt]:[vɛrtə] 'value(s)' points to Pre-German **vert*. The historically attested wordform is the OHG *werd*, which was replaced by the final devoiced wordform (cf. Ringe 2003:247).

Consequently, an alternative view on IR considers only INTERNAL RECONSTRUCTIONS OF SOUND CHANGES possible. The internally reconstructed sound change *t* > *s* before *i* from the Greek word *ἀμβροσία ambrosía* only indicates that *s* goes back to *t* here. The method does not allow for further inferences about the rest of the wordform. This view also underlies this thesis although in the literature, wordform reconstruction is often mentioned as a goal of IR (e.g., Bauer 2009:18).

3.6.3 Language Typology and Reconstructability

Table 3.H. Development of the singular paradigm of Old Irish fer 'man' (cf. Sims-Williams 1998:364).

	Proto-Celtic	**Old Irish**
Nominative	**wir-os*	*fer* [-r]
Genitive	**wir-ī*	*fir*^{L} [-r^{j}]
Dative	**wir-ūi*	*fiur* [-r^{u}]
Accusative	**wir-om*	*fer*^{N} [-r]

The basic hypothesis of IR, in its structuralist definition, requires the existence of allomorphy in the examined language to initiate the reconstruction process. The amount of allomorphy may depend not only on the language's general regularization tendencies but also on its typological structure. Strong allomorphic tendencies are considered to be a feature of inflectional languages, while typically agglutinative languages form little allomorphy (Selig 1992:87), and isolated languages (Hasil 2008:135) no allomorphy. Sound changes and the associated weakening of morpheme boundaries are not an obstacle to language with fusional linguistic constructions and can thus prevent the tendency to regularize. The Old Irish singular inflection of *fer* 'man' can serve as an example (Tab. 3.H). We recognize here, on the one hand, vowel allomorphy, which came about through the transfer of phonetic information from the former following morpheme. On the other hand, *ī* and *ū* caused a palatalization and velarization of r, respectively. This palatalization can still be understood as a morpheme in this context although its *signifiant* is realized at the same time as the previous *signifiant* of *fir*. Such an interplay of morpheme and sound can be attribute to the internal reconstructability of languages (Hoenigswald 1944:86).

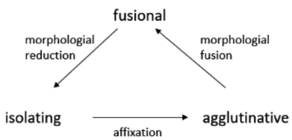

Figure 3.6: Typological circle of languages.

Polysynthetic languages, as was assumed at the early stage of IR (Chafe 1959:484), offer little material for IR because they tend to preserve morpheme boundaries. If a sound change occurs at this position, a strong tendency toward regularization is to be expected. In particular, agglutinative languages tend to have a simple and regular morphological structure and are therefore not considered very suitable for IR. Anttila (1973) was one of the first to address the issue of IR and agglutinative languages, using Finnish as an example. He saw a major difference in the relation of language to the diachronic projection: While irregularities in fusional languages indicate diachrony, alternations in languages with perfect agglutination are more likely to occur when an innovation becomes productive and spread; the agglutinative language "looks into the future" rather than into the past (Anttila 1973:317). Perfect agglutination, however, is rarely found in natural languages. In Anttila's study of Finnish, consonants offered several starting points for IR, including consonant stages (Anttila 1973:319–324 and 339). Vocalic pre-phonemes and sound changes were less reconstructable. The results of this IR study could be supported "[o]n the whole" with comparative evidence (Anttila 1973:348). However, Anttila (1973:349) also qualified that all reconstructions contained errors. Among these errors are alternations caused by loanwords. In these cases, "internal reconstruction was derailed" (Anttila 1973:349).

The theoretical starting point of IR in agglutinative languages is also allomorphy (i.e., the language's fusional part). If this part is missing, as in purely agglutinative languages, the possibility of morphophonemic IR is consequently missing. Conversely, IR dissolves the fusional character of a language, which led Anttila (1973:346) to think: "It would of course be typologically very interesting to carry the method to its dead end. We would easily end up with perfect agglutination." The fact that IR creates a regular and idealized preface is a well-known point of criticism that has been addressed a number of times (see Sect. 6.2.3). From a purely theoretical point of view, this development actually corresponds to the *typological circle* (also *morphological cycle*; see Fig. 3.6) based on the work of Humboldt (1822). A more recent definition was given by Dixon (1994:182–183): "a fusional language can develop into one of the isolating type, an isolating language can became agglutinative, an agglutinative language may move towards a fusional profile, and so on." The typological circle is considered as a simplified picture, which has therefore been criticized several times (e.g., Iguarta 2015) but nevertheless describes a trend in language evolution. Within Fig. 3.6, allomorphy arises during the process "morphological fusion." IR is only possible for those parts of the language that are in this phase of development. Parts that are in another phase are not taken into account by IR, so we would inevitably end up in a "perfect agglutination." To better grasp further development stages of the typological circle through IR, other basic hypotheses are required.

Agglutinative basic hypothesis | The agglutinative basic hypothesis assumes that bound morphemes originate from a free morpheme. This hypothesis projects back a bound morpheme onto a lexeme of similar meaning. A reconstruction in the sense of a phonetic restoration of the original form or a sound change does not take place by itself. If we apply the agglutinative basic hypothesis to the Gothic wordform *sniumidedum* 'we hurried (1^{st} pl. prt. ind. act.)', an IR could proceed, as seen in Tab. 3.I.

Table 3.I. IR on the Goth. sniumidedum as example of an agglutinative IR. All morphemes are analyzed as Pre-Gothic free morphemes.

	Morphological Analysis			
Morpheme:	*snium*	*-i*	*-ded*	*-um*
Function:	**snium-s* ‚fast'	causative or factitive	prt.	1. pl.
Reconstruction:		**i-* ‚factitive or causative verb or noun, e.g., to make'	**ded-* ‚preterital auxiliary verb or particle'	**um* ‚1. pl.'

In the next step, the presence of cognates can be postulated to enable a phonetic reconstruction. The procedure then corresponds to that of traditional IR, but the reconstruction is to be classified as less likely due to its uncertain cognate status. The form **um* could be interpreted as an old nominative of the suppletive paradigm for the personal pronoun of the first-person plural. These Pre-Gothic reconstructions would not be correct, but it is not uncommon to assume that the verbal suffix has such an origin (e.g., Kapović 2017:83). Similarly, the morpheme *-ded-* could indeed be derived from an auxiliary verb (cp. OHG *tâten*), but other explanatory models are possible here as well (see Braune and Heidermanns 2004:155). However, a verb (or particle) **i* 'make' cannot be easily accepted. Etymologically, the Gothic morpheme can be traced back to different origins (Braune and Heidermanns 2004:158).

This basic hypothesis has a similar problem to that of the HEA, so it does not need to hold true for all bound morphemes. Morphemes can also arise from re-analyses. In addition, reinterpretations or mergers may produce a semantic or phonetic innovation that can no longer be restored internally. This method shows similarities to Cratylic reconstruction (see Sect. 5.1) but differs from it in its exclusive application to morphemes and its preservation of morpheme boundaries. A similar situation is deemed for the basic hypothesis of isolating languages.

Isolating basic hypothesis | The isolating basic hypothesis assumes that one-morpheme lexemes originate from multi-morphemic lexemes. The additionally assumed morphemes may have been lost or completely fused with the preserved morpheme. The first case cannot be reconstructed; in the second case, the morphemes may have left reflexes that may be reconstructed. This is true when, for instance, a set of semantically different but functionally similar words share a phonetic commonality.

For example, factitive verbs in German often contain an *umlaut*. If the root word ends in a fricative, the derived verb additionally shows an affricate (Tab. 3.J). These cases can be reconstructed internally by what is referred to as *pattern reconstruction* (see Sect. 4.2.3). In practice, the fused morpheme is often difficult to reconstruct since all that remains of the original morpheme is a small reflex, which allows little more to be said than that it once existed.

Table 3.J. Samples of "fused morphemes." The derived verbs contain umlaut due to the Germanic suffix *-j-an.

Root Word	Derived Verb
schwarz ‚black'	*schwärz-en* ‚blacken'
offen ‚open'	*öffn-en* ‚to open'
Hass ‚hate'	*hetz-en* ‚to agitate'

4. Methods of Internal Reconstruction

4.1 Reconstructability of Sound Change Types

Before discussing the methods of IR proposed and applied in the research so far, the issue of reconstructability should be recapitulated in this section. The question of the methods' goals depends on the question of reconstructability (i.e., what can be reconstructed INTERNALLY). Here, as has been implicitly done in the literature, the concept is largely restricted to the phonological level. Since a restriction to sound change is also useful for automating IR for practical reasons, this restriction will continue to be applied in this thesis.

Different types of sound change cause different changes in phonotactics. This results in different degrees of internal reconstructability for sound changes. A fundamental differentiation of sound changes depends on the question whether a phonemic change results with it (i.e., whether there is a so-called *phonemic split*). Phonemic splits occur when an allophone of a phoneme becomes an independent phoneme or merges with another.

Sound changes with phonemic changes can be categorized as a *primary* or *secondary split* according to Hoenigswald (1974:190–199). In a primary split, the splitting phoneme itself is affected by the sound change (e.g., *k* > *c* \ _V[+palatal]). In a secondary split, the phoneme splits only due to merging or loss of the conditioned environment, such as in many *umlaut* phenomena. The starting point of IR is the result of the sound change and the method goes backwards from there. According to Hoenigswald (1992:28), the following rule applies: “a ‘regular’ alternation points to a primary split, under (at least partly) stable conditions; an ‘irregular’, morphemically-conditioned alternation points to a secondary split.” The following categorization of sound change types is based on the typology of Hoenigswald (1946). Fig. 4.1 gives an overview of the different kinds of non-phonemic changes and primary splits.

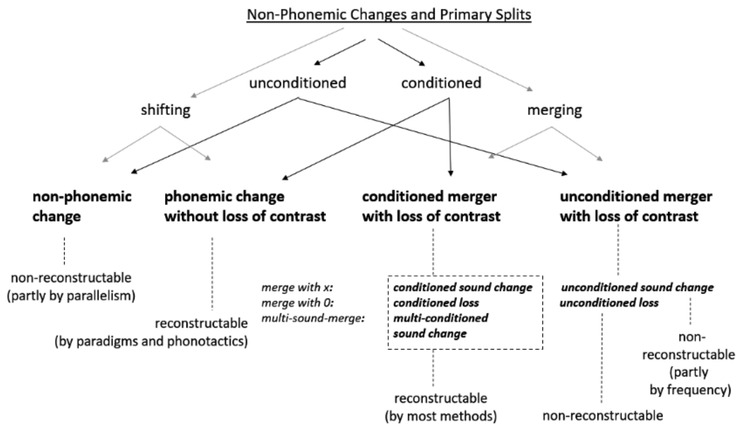

Figure 4.1: Summary of the reconstructability of primary splits.

4.1.1 Non-Phonemic Change

Hoenigswald (1946:138) defined "non-phonemic" as a type of sound change that only changes the articulation of the phoneme without any other changes of the phoneme — be it in an allophonic, morphophonemic, or phonotactic respect. These include mainly context-free sound shifts (*shifting*), such as the plosive rotation in Armenian or parts of the Germanic Grimm's law. Since shifting does not leave any changes in phonotactics, there are almost no opportunities to conduct IR. However, reconstructions of non-phonemic change are possible via the principle of parallelism. Latta (1978:49–50) cited the *anlaut* nasalization of voiceless plosives in Javanese as an example of this (Tab. 4.A).

Table 4.A. In Javenese, transitive verbs are derived from nouns and intransitive verbs through nasalization of the first sound (cf. Latta 1978:49–50).

Root Word	Transitive Verb
*p*ayung 'umbrella'	*m*ayung 'to cover so. with an umbrella'
*t*iliʔ 'to look at'	*n*iliʔ 'to look for so.'
*t'*abut 'expression'	*n'*abut 'to express sth.'
*s*urung 'to shove off'	*n'*urung 'to shove sth.'
*k*ukup 'to take in hand'	*ng*ukup 'to take sth. in hand'

The Javanese nasalization is homorganic: the plosive is replaced by a nasal consonant at the same place of articulation. Consequently, no palatal nasal would be expected for the dental sound *s* in *surung* > *n'urung*. Since the morphophonemic rule otherwise concerns only plosives, a previous **t'* or **č* can be assumed for *s* (cp. Latta 1978:50). King (1971:213) cited the change Proto-Indo-European **ei* > Germ. **ī*, which continues to appear as diphthong in *ablaut* paradigms. Both methods of reconstruction are forms of *pattern reconstruction* (see Sect. 4.2.3).

4.1.2 Primary Split

4.1.2.1 Phonemic Change Without Loss of Contrast

This category includes sound changes in which some occurrences of a phoneme A are reassigned to another phoneme B that did not previously occur in that environment. This applies to the exceptions of Grimm's law after voiceless consonants. Since in PIE voiceless plosives were followed by a voiceless plosive or **s*, in this position, there was only a Pre-Germ. **t*, not **d*. In this environment, the Proto-Indo-European **t* merged with the phoneme Germ. **t* (< Proto-Indo-European **d*), while Proto-Indo-European **t* in other environments was shifted to **þ*. According to Hoenigswald (1946:139), this change is still reconstructable by the alternation of Gothic participles *salbo-þ-s* 'anointed' (inf. *salbo-n*) and *haf-t-s* 'bound, afflicted'[27].

4.1.2.2 Conditioned Merger with Loss of Contrast

If the split sound falls to a phoneme that has already occurred in this position, the merger is called "with loss of contrast." Three special cases of this can be distinguished.

Conditioned Sound Change | This first type of these sound changes corresponds to the "classical" conditioned sound change. The result is a merged phoneme depending on the phonetic environment, while the phonemes in other environments remain distinctive. In Hoenigswald's (1946:140) terminology, the "old sound" is now "restricted," but the new one is "free." This type of sound change is reconstructable in paradigms using morphological-distributional methods (see Sect. 4.2.1 and 4.2.2). Note, however, that these cases are *synchronically* indistinguishable from merger *without loss of contrast*, which constitutes a potential source of mis-reconstruction.

[27] The Goth. *hafts* is a frozen participle of an unattested verb **haf-* (< PIE **keh₂p-*; cf. Lat. *capere* 'to take'). Nevertheless, Hoenigswald's reconstruction is questionable (see Miranda 1975:300).

Since the free sound also occurs both with and without alternation in the same environment, it is "indeterminate" (Hoenigswald 1965:102) for non-alternating paradigms, whether the sound in those environments goes back to the "old" sound or not. For instance, it cannot be said for German *und* [ʊnt] whether the final *t* is original or goes back to a final devoiced *d*. Nevertheless, indeterminate cases are relevant for IR when analogical levelling has covered a sound change in paradigms because these cases usually have no analogical pressure.

Conditioned Loss | A sound disappearing in certain environments (*conditioned loss*) does not make any difference in terms of its IR. This sound change may still be reconstructable via paradigms showing alternations with Ø-morpheme (see Hoenigswald 1946:140).

Multi-Conditioned Sound Change | It may happen that one sound merges with different sounds in different positions. This case does not differ from the others. However, a special case occurs if a sound disappears from the phoneme system by distributing its phones to several other phonemes. An example is the Proto-Indo-European laryngeals in Sanskrit, which disappeared or were reflected by *i*, *a,* or a vowel lengthening. In paradigms with the old sound, alternances can arise, but this "alternation will not be compulsory in point of phonemic occurrence" (Hoenigswald 1946:139). The "new" phones of /i/ and /a/ normally overlap in their distributions and therefore show the alternation exclusively in single paradigms.

4.1.2.3 Unconditioned Merger with Loss of Contrast

Unconditioned Sound Change | If all sounds of one phoneme merges with those of another phoneme, there is an "unconditioned merging." This includes, for example, the merging of Proto-Indo-European *d^h and **d* in Slavonic. These cases do not provide a starting point for morphologically based IR. For that reason, statistical approaches were proposed (see Sect. 4.2.2.2) because the "process leaves only inconclusive traces in the form of an occasional statistical preponderance of the surviving phoneme" (Hoenigswald 1946:136).

Nevertheless, different morphologically based IR of this type have been presented in the literature. Ringe (2003:248) cited the merging of Pre-Greek **aː* and **eː*, which can be seen by the fact that short *a* alternates with *eː* (among others in the plural endings of the first declension), whereas old *ɛː* alternates with short *ɛ*. In fact, this reconstruction is only possible due to a preceding sound change causing the alternation of short and long *a*, which makes it a case of a "secondary split." A similar case is the reconstruction of the merging of Proto-Indo-European **e, *o, *a* to Sanskrit *a*, which can be reconstructed synchronically via the preceding sound change **k* > *c* before old **e* (see Bynon 1977:98).

Unconditioned Loss | If a sound merges unconditionally with the Ø-morpheme, there are no possibilities of internal reconstruction. This case applies, for example, to the Latin *h*, which disappears in most Romance languages without any relicts.

4.1.3 Secondary Split

In addition to the type of sound change, reconstructability may be limited by secondary developments that cause the loss of an alternation's automaticity. Hoenigswald (1944:82–83 and 1946:140–142) distinguishes three cases of secondary splits:

- Loanwords
- Analogy or neologisms
- Sound changes that affect other sounds (usually those of the conditions)

The largest and probably most dominant factor are secondary and subsequent sound changes (Sect. 4.1.2.1) that cover the earlier sound changes. Its influence on to the reconstructability depends individually on various factors, such as the intersection of the affected wordforms, the reconstructability of the subsequent sound change, or the type of the subsequent sound change. In complementary allophony, such as German [x] (after back vowel) and [ç], allophones only occur in different positions (see Fig. 4.2).

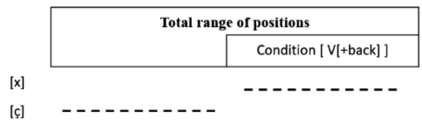

Figure 4.2: Illustration of complementary allophony using the German sounds [x] and [ç]. The scheme is freely adapted from Hoenigswald (1944).

Due to secondary factors (such as neologisms like *Frauchen* 'mistress [of a dog]'), a weak distributional overlap between both phones occurs, which can produce minimal pairs and thus influence the phoneme status. The new distribution is not completely complementary anymore but can be considered "quasi-complementary" (Hoenigswald 1944:85). These sound changes can still be reconstructed internally as long as the distributional overlap is not yet well advanced. However, there is a possibility that an alternation might be misinterpreted as "allophony" if the invariant is not attested purely by change (see Mehendale 1963:42).

The reconstructability of "secondary splits" is mostly seen very critically. For secondary split reconstructions, a new assumption about the "phonetic naturalness" of sound changes and uncertain sound change sequences must be formulated, which makes the IR of secondary split "much more speculative than in the [other, A.B.] cases" (Ringe 2003:253). Ringe (2003:253–255) explains these additional assumptions using the example of alternation *f:v* in English:

Table 4.B. The English alternation *f:v* is present in few paradigms (first column) and missing in most other cases (Ringe 2003:253–255).

f : v		*f : f*		*v : v*	
leaf	/liy*f*/ : /liy*v*z/	reef	/riy*f*/ : /riy*f*s/	sleeve	/sliy*v*/ : /sliy*v*z/
knife	/nay*f*/ : /nay*v*z/	fife	/fay*f*/ : /fay*f*s/	five	/fay*v*/ : /fay*v*z/
loaf	/low*f*/ : /low*v*z/	oaf	/ow*f*/ : /ow*f*s/	stove	/stow*v*/ : /stow*v*z/

Columns 2 and 3 in Tab. 4.B make it clear that the alternation is not automatic in modern English and has been obscured by a secondary development. This development requires additional assumptions. As a first suggestion, Ringe (2003:254) proposes the assumption of a separate proto-phoneme (e.g., **β*) for all three groups and a secondary development of **β* > *f* / _# and **β* > *v* / *else*. However, this development can be rejected as "phonetic[ally] unnatural" since the implicit sound change **βs#* > *vz#* would be unusual (Ringe ibid.). To compensate for this problem, one would need another assumption, such as a vowel loss in the first group (**βVs* > **vVs* > *vz;* ibid.). The actual cause for this alternation is merely one of many possibilities that cannot be directly inferred by IR. Only in a few conceivable cases, therefore, a wordform can actually be reconstructed through this method. In most cases, it only enables us "to infer the existence of distinguishing morphological contexts in these cases" (Fox 1995:164).

4.1.3.1 Secondary Re-Arrangement Induced by a Primary Loss of Contrast

De-Restriction by a New Sound Change | If a phoneme is "restricted" in its distribution, this gap can be filled by a new sound change. Pre-Greek **s* disappeared in intervocalic positions but reappeared in this position due to the more recent sound change **ti* > *si*, resolving the distributional restriction of *s*. According to Hoenigswald (1946:141), instances of a new sound change are mostly only reconstructable "where zero is a party to a non-compulsory alternation." However, Mehendale (1960:102–103) saw a possibility for reconstruction in some circumstances. Assuming /x/ has merged with /y/ after the first sound change, [x] is a restricted allophone of /y/. In the following sound change, /z/ becomes [x] in positions where no [x] has been found previously. Now /x/ is "less" restricted and /x/ and /y/ are no longer compulsory. /z/ alternates with /x/ (but not vice versa), which is internally reconstructable (Mehendale 1960:102). Through subsequent IR based on the first reconstruction, the first sound change can also be reconstructed, even without the involvement of "zero" (see Fig. 4.3).

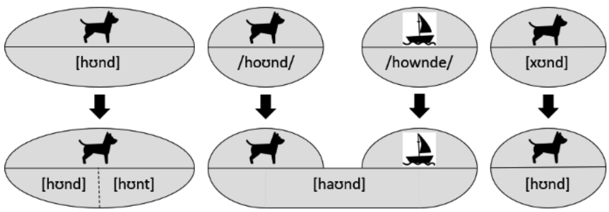

Figure 4.3: Illustration of Mehendale's example: /d/ and /t/ merge in (1) with an automatic alternation (2). A second sound change unlocks the restricted distribution of [t] (3).

Loss of Condition | Let us assume a complementary allophony similar to that of Fig. 4.2 and an omission of the condition by a second change splitting the allophones into two different phonemes. Concrete examples would be the German *umlaut* or the English *th*-sounds. The *umlauts* originally appeared as allophones if an *i* or *j* was part of the subsequent syllable. After its loss and neutralization to *ə*, respectively, the condition of the phonological rule failed, and the *umlaut* became phonemes.

If it is possible to reconstruct the loss of the condition as such, the sound change can be reconstructed in principle, but this is rather unlikely in practice. Some Welsh wordforms show an alternation of *h*- and *s*- in the initial position (e.g., *hil* 'issue, progeny, offspring' and *sil* 'issue, seedling, spawn'), suggesting a Pre-Welsh sound change *s*- > *h*-. However, it seems to be unlikely that the correct reason for this alternation "would ever be achieved by the synchronic method" (Fowkes 1950:143).

Transferring an Allophone into a Paradigm | After a sound change, an allophone may be transferred into a paradigm to create a contrasting position. As an example, Hoenigswald (1946:141) cites the slightly diphthongal sound [æ·] in "General Eastern American English," which occurs before certain or all voiced consonants in syllables' final position instead of [æ]: *to pad* [pǽ·d] in contrast to the noun *padding* [pǽdıŋ]. The gerund *padding* is pronounced as [pǽ·dıŋ] because "the verb form at one time differed from the noun by showing features of open juncture […] which was later lost, thus putting [æ] and [æ·] into a contrasting position" (ibid.).

4.1.3.2 Analogical Levelling

Secondary analogical levelling can interrupt an automatic alternation. The reconstructability of sound changes is decisively disrupted if all paradigm wordforms and derivations are completely equalized. The final *-r* of the Latin *honor* 'honor' is adapted to *honor-is* (gen.sg.) < *honos-is*. The paradigm is analogically levelled, but the original sound can still be derived by *hones-tus* 'honorable'. However, an alternation that is only attested by one allomorph pair (e.g., Latin *honor* and *hones-tus*) can also be explained by suppletion.

Analogical levelling can also create paradigm splitting, if two paradigms that continued to exist as *near-synonyms* in *doublet paradigms* are created on the basis of wordforms of the old paradigm. According to Hoenigswald (1965:111), the more productive paradigm may indicate which sound is older (e.g., the more productive Latin paradigm of *honor* compared to the archaic paradigm of *honos*). The problems with this conclusion are discussed in Sect. 4.3.1.

4.1.3.3 Borrowing

An allophone can also dissolve its distributional restriction by borrowing. An example of this is the Russian /f/, which alternates with [v] in its native vocabulary. It was only through loanwords with *f* that it also appeared in the same positions as [v] and thus became an own phoneme.

4.2 Types of Phonetic Internal Reconstruction

The traditional approach of IR, which examines alternations in allomorphs, therefore pursues a morphological approach. This constitutes the most common method of IR, such that it is sometimes referred to as the "normal IR" (cf. Boretzky 1975:56).

Other approaches than this have been discussed and proposed across time, sometimes with strongly different designs and aims. Their admissibility or their meaning were also repeatedly questioned. In the view of Chafe (1959:478), the comparison of internal cognates seems to be "the only conceivable basis for internal reconstruction." This limitation, however, unnecessarily narrowed the potential of IR and forced the method to stagnate in the history of science.

4.2.1 Morphophonemic Internal Reconstruction

The starting point of morphological IR is with the allomorphs of a morpheme. The relationship between two phonetically different allomorphs of the same morpheme is called "morphophonemic alternation" (Wells 1949:100). Classical morphological IR is therefore also called *morphophonemic IR*. Excluding suppletion, a pre-historical proto-form is assumed (*hypothesis of etymological allomorphy*) to be the origin of recent allomorphs. Whether these morphs are grammatical, lexical, bound, or free hardly matters. Marchand (1956:246) refers to these morph pairs using the term "internal cognates," which generally refers to the methodological similarities of internal and comparative reconstruction (cf. Mehendale 1963:41, Lass 1975:5, Fox 1995:154, Sihler 2000:150). Alternative terms are "tautolinguistic cognates" (Lass 1997:232) or "cognate morphs" (Chafe 1959:479). Then, the historical linguist examines all internal cognates available in the language stage and derives reconstructed *base forms* (a term coined by Hockett 1958:463) and sound laws from these internal cognates.

For a sound change to produce an allomorphy, the morphemes must meet two conditions: The condition of the sound change must be outside of the morpheme, and the morpheme must occur in the environment both with and without the condition (cf. Boretzky 1975:51). If these conditions are not met, the morpheme is an *indeterminate* morpheme. Such morphemes play a role in morphological-distributional methods, as these continue to allow a statement about the conditions of a sound change. Although the condition must lie outside of the morpheme to be able to create any alternation, it was stated at the same time that morpheme boundaries are not suitable for the determination of conditions of sound changes (see Zeps 1969:150). New morphemes may be attached to the stems, which can obscure the actual cause of the sound change. For instance, Grimm's law did not occur after fricatives, so the sound sequence *-fþ-* did not result and this exception of the sound shift could be reconstructed internally. Nevertheless, new formations such as the Goth. *afþliuhan* 'to flee away' have been created from *af* 'from, away' and *þliuhan* 'flee'. By *excluding sound sequences at morpheme boundaries*, these cases can be ignored.

4.2.1.1 Initial Data and Cognate Pairing

At the beginning of the inductive process, the Comparative Method sets up a series of potential cognates, which are confirmed or refuted later on. Starting points for connecting such potential cognates are usually based on the phonetically and semantically similarity of both words. Semantics serves as a pre-selecting element and can reduce randomly similar pairs.

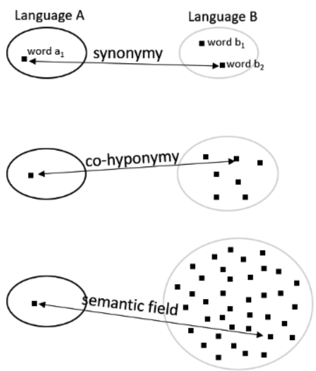

Figure 4.4: Words of a cognate pair can be synonyms, co-hyponyms, or they belong to the same semantic field. The greater the number of potential cognate words in Language B, the more unlikely is the cognancy.

The question of predefining cognates in IR is rarely addressed in the literature. Already in Hermann (1907), wordforms — semantically related as well as semantically distant words (e.g., ancient Greek *ἔνδον éndon* 'inside' and *δόμος dómos* 'house') — were connected without further explanations. The fact that a connection, such as the one between "inside" and "house" in ancient Greek, would barely be derivable from a purely internal point of view or would only be possible with additional speculations is not addressed. Thus, this constitutes a potential weakness in the method. If the cognate pair *ἔνδον* and *δόμος* is a legitimate pair, random pairs cannot be prevented. Pairs are judged permissible by most linguists (e.g., Boretzky 1975:52) when their outer forms bear some resemblance to each other and there is also some similarity in their functions and meanings. In fact, however, the type of cognate determination is an important factor in identifying the degree of probability of a sound change. Synonymy between two potential cognates is a stronger indicator for cognancy than co-hyponymy because the number of co-hyponyms is much larger, increasing the likelihood of coincidence (cp. Fig. 4.4). Synonymy is also suitable for predefining internal cognates in IR. Different criteria may be relevant for pairing two cognates. At the morphological level, for example, they should belong to the same paradigm, or they should be derivations. In this thesis, three criteria of predefining cognates are assumed.

In PARADIGMATIC PAIRING, allomorphs are connected by a paradigm and mostly separated by different attached inflectional morphemes (e.g., *sing-s* ~ *sang-Ø*). The paradigm can be permeated by "real" suppletion. In the case of DERIVATIONAL PAIRING, allomorphs are connected by semantics and separated by different productive derivational morphemes (e.g., *to sing* ~ *the singer*). At this level, fare more suppletive forms may occur (e.g., *to cook* ~ *the cook* instead of †*cooker*). The third connecting criterion is a purely SEMANTIC PAIRING. Its cognates are still held by the same semantic word field but can no longer be derived from each other synchronically (e.g., *to sing* ~ *song*). Internal reconstructions based on this pairing type assume an unproductive paradigmatic (e.g., *to cleave* ~ *cloven*) or derivational pairing (e.g., *stink* ~ *stench*). The basic hypothesis of etymological allomorphy is difficult to maintain here since it is not supported by additional information (linguistic competence) and is based only upon semantic similarity. In accordance with these three types of pairing, this thesis distinguishes between three sub-methods of IR:

- Paradigmatic approach
- Derivational approach
- Semantic approach

In the literature, not only morphemes but also grammatical bound morphemes, such as the English plural morpheme /-s/ ~ /-z/ ~ /-əz/ (*rat-s* ~ *key-s* ~ *hous-es*) or the participial suffix /-d/ ~ /-n/ (*hear-d* ~ *see-n*), are used for IR. While the example of the plural morpheme legitimizes the use of affixes as starting material, the example of the participial suffix reveals the high degree of suppletion observed in grammatical morphemes. Therefore, no better overall result can be expected by the additional inclusion of affixes. This is truer for alternations of syntactic constructions, such as the word order *V2* and *VL* in German (cf. Hock 2013:12–14).

Chafe's Special Cases of Internal Reconstruction | These three approaches can be supplemented by "two specialized guises" according to Chafe (1959:479), which include variants of sandhi and stylistic formulations. Chafe defined these as alternative forms "between which the only semantic difference is the same as between another such pair, while the morphemic content of the first pair has nothing in common with that of the other" (ibid.). Both variants can be added as SANDHI-SYNTACTIC PAIRING and STYLISTIC PAIRING to the types presented above. However, the morphophonemic method, unlike the Comparative Method, does not actually compare words but only individual allomorphs (ibid.). In the case of stylistic variants, there is a "real" synchronic cognancy: the words are cognates, not just containing the same morpheme. This is associated with a modification of the basic hypothesis, which means that it actually constitutes a new reconstruction method. However, since the further procedure corresponds to morphological IR, a separate presentation is renounced here. To determine the original form, the stylistic marked variant must be identified. Givón (1990:947) defines three characteristics of markedness that tend to make it possible to determine the older variant:

- Structural complexity: the marked form tends to be "structurally" more complex
- Frequency distribution: the marked form tends to be more uncommon
- Cognitive complexity: the marked form tends to be "cognitively" more complex

One problem with these characterizations is that a stylistic variant can become an unmarked variant over time and vice versa.

4.2.1.2 Determining Sound Correspondences

If phonetic alternation repeatedly occurs in allomorphs, a phenomenon called "recurrent phoneme correspondences" (Chafe 1959:479) arises and these alternations may be candidates for a diachronic sound change. To sort out a multitude of false potential "correspondences," a filter of "dissimilar" phonetic pairs can be used (cf. Hockett 1958:467). Chafe (1959:479) rates a sound pair as "similar" if the words share at least one distinctive "member."

For instance, sound pairs may include *t/d*, *d/d*, *t/t*, and *d/0*. In the next step, different phonetic environments can be determined for the sound pair. While *t/t* appears in all positions in German, the pair *t/d* occurs only in final positions and before voiceless consonants. The pair *d/d* does not appear in exactly this position and can thus be considered a COMPLEMENTARY PAIR. A "complementary distribution" is called the distribution of two environments A/B and C/D if they have nothing in common *or* if "A and C have something in common but B and D do not (or vice versa)" (Chafe 1959:480).

Figure 4.5: Own illustration of Hoenigswald's rule: [X] is complementary distributed to [A] and [B] and, therefore, is assigned to the phonetically closer phoneme /A/. A sound change [A] > [C] causes a new assignment to [B] that is phonetically closer afterwards.

Another concept for determining sound correspondences is explained by Hoenigswald (1946:138,1992:23), who sees phonetic similarity as a basic principle of allophony. If a sound [X] occurs in a complementary environment to [A] and [B], "[X] will naturally be assigned to the same phoneme" (Hoenigswald 1946:138). If the sound [A] changes, so that [B] in phonetically closer to [X], the sound [X] changes to the phoneme [B] (ibid.; see Fig. 4.5). The German sound [ɐ], which is an allophone of [r] and is likewise distributed complementary to [h] and [j], may serve as an example. If [ɐ] and [r] no longer alternate in the allomorphs, [ɐ] should become an allophone of /j/ according to Hoenigswald's rule. Nevertheless, in German, [h] and [ŋ] are not considered allophones either, despite their complementary distribution. Whether this rule applies universally remains under consideration.

4.2.1.3 False Sound Correspondences

Identified sound correspondences need not come from a diachronic sound change. This is true for suppletion, analogy, and irregular changes. While the basic hypothesis is already incorrect for suppletive forms, it still applies in the other cases. Irregular sound changes cannot be identified as such, but they are also problematic when using the Comparative Method (see Chafe 1959:480 and Latta 1978:24). However, the Comparative Method offers more possibilities for identifying such cases.

Analogies normally regularize paradigms. Irregularities caused by analogies usually arise through inter-paradigmatic analogies, not by intra-paradigmatic ones. This is therefore a problem that occurs more frequently in the derivational approach and is limited in the paradigmatic approach to cases where one paradigm adopted irregularities from other paradigms. Latta (1978:23) mentions borrowing from another dialect or related languages as further cases, without giving examples of borrowed wordforms in inflectional paradigms. These cases are also more likely to be found in the derivational approach.

4.2.1.4 Identifying the Direction of Sound Changes

The determination of the direction of a sound change is the next step after determining a sound pair *A-B* (i.e., *A* > *B* or *B* > *A*). By this step, the pre-phoneme of both sounds has to be identified. As in comparative wordform reconstruction, the transition from the sound correspondences to the proto-forms can be observed through inferring methods.

In the simplest case, the direction of sound change can be determined to be phonetically impossible or unlikely (PHONETIC PLAUSIBILITY). For example, *s* > *h* is phonetically more plausible than *h* > *s* and the direction *Vn* > *Ṽ* is more frequent than *Ṽ* > *Vn*. The use of universals or empirically determined probabilities for sound changes may be applied in those cases. However, there is, on the one hand, the uncertainty of actually having a universal rule and, on the other hand, the possibility of an intervening sound change. For instance, if the conditional sound change **ç* > *s* was followed by an unconditional sound change **ç* > *h*: the resulting complementary pair *s*:*h* would infer a wrong direction.

Once the condition of a sound change is determined, a conclusion about the pre-phoneme can also be drawn via THE ARTICULATORY SIMILARITY OF A SOUND PAIR AND ITS CONDITION (see Boretzky 1975:52). Since different phonemes must differ phonetically in at least one distinctive feature, it can be concluded that a phonetic distinction is necessary to make a *phonemic split* possible. This new feature of the newly created allophone can (almost) only be inherited from the phonetic environment. An alternation *m / _p ~ n / _{-p}* indicates an assimilation process and thus the sequence *n* > *m*. However, the possibility of negative conditions (*m* > *n* except before labials), as well as dissimilating phenomena, cannot be excluded in this way. The question of syncope or anaptyxis is also readily made to depend on the surrounding consonants or other phonetic factors of the environment. For Chafe (1959:482, fn. 15), epenthesis is plausible only if the environments "are definable in strictly phonological terms."

Table 4.C. Scheme for final devoicing of *g:k* according to Miranda (1975:296).

	_V	_#
g	+	-
k	+	+

A popular method for determining the direction of a sound change, called the "standard approach" by Miranda (1975:296), uses PHONETIC DISTRIBUTIONS and sees the "restricted sound" as the original pre-phoneme. Tab. 4.C illustrates the German final devoicing as in *Tag* [taːk] 'day' and *Tage* [taːgə] 'days'.

In this case, *g* is the *restricted sound* and, according to Miranda's rule, the direction *g* > *k* can be determined. If both sounds are allophones of the same phoneme, they are distributed in a complementary way, and the scheme shows two empty fields in different rows and columns. In this case, a determination of the phonetic pair is possible, but the direction of the sound change can no longer be determined. However, even the first case does not say much about the diachronic development, as already noted by Mehendale (1960:101–102) and Miranda (1975:299). A misinterpretation could arise if the phonemes or the conditions coincide. The example in Tab. 4.D shows the distribution of Sanskrit *c* and *k* before and after the merging of *e* and *o* to *a*.

Table 4.D. Distribution of Skt. c and k before (1) and after (2) the merger of *e and *o > a (cp. Miranda 1975:301).

(1)	e	o	{#,s}
č	+	-	-
k	-	+	+

(2)	_a	_{#,s}
č	+	-
k	+	+

From this case, Miranda concludes that many phonemic splits are reconstructable. Only if "the later developments are complete mergers or some such changes that leave no trace […] there is no way to detect" the involved sound change pattern (Miranda 1975:302). The approach is only correct if one can assume a conditioned merger, but these may hardly be distinguishable from "former" allophonic distributions.

4.2.1.5 Multi-Sound Alternations

If one sound alternates with several other sounds and the others do not alternate with each other, the former sound can be determined as the pre-phoneme (see Anttila

1973:322). The question then turns to the resulting sound: Is the multi-sound alternation the result of one or more sound changes? An example of a multi-sound alternation is found in Finnish paradigms, where there is a plosive alternation in open versus closed syllables (Tab. 4.E).

Table 4.E. Finnish plosive alternations in open and closed syllables (Anttila 1973:322).

$/p_1/$	p:p	$/t_1/$	t:t	$/k_1/$	k:k
/b/	b:b p:v	/d/	d:d t:d	/g/	g:g k:0
$/p_2/$	p:m p:0	$/t_2/$	(s-)t-d t-l t-r t-n t-0	$/k_2/$	k-v k-j k-ŋ

The basic hypothesis assumes an original sound **X*. Only the sounds of the open syllables can be considered the pre-sounds in Tab. 4.E since the forms of the closed syllable can only be derived from these sounds, and not vice versa (cf. Anttila ibid.). The reverse case (e.g., **X = v*) is conceivable only in theory with *v > p* in open syllables and *v > m/0* in conditional closed syllables. Anttila (1973:323) posited a fricative intermediate stage from the fact that the resulting sounds all have a “continuate articulation and voice” but also for historical and dialectal reasons (Fig. 4.6). Whether the intermediate stage is reconstructable without “external” information remains questionable. The parallel structure of the plosives leads to the assumption that there must be a parallel development, but this assumption is made through a highly simplified picture; other possible intermediate stages cannot be excluded.

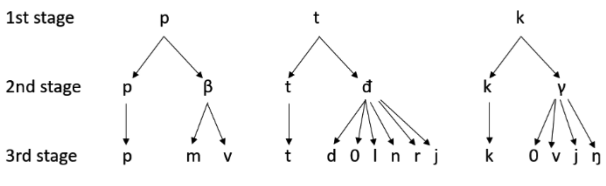

Figure 4.6: Sample of a multi-sound alternation: the development of the Finnish plosive alternation (Anttila 1973:323).

4.2.1.6 Reconstruction of Sound Change

With the determination of the pre-phoneme, an internal wordform reconstruction is completed by substituting the synchronic sound in the wordform with the pre-phoneme. The sound change reconstruction continues the procedure for determining the CONDITION and for ABSTRACTING concrete sounds. Hoenigswald (1944) was the first to attempt to describe the formal foundations for applying IR to sound change. From the phoneme's point of view, the differentiation of diachronic sound change and synchronic phonological alternation does not play too much of a role since the latter is only regarded as a more recent sound change. If the sound change or alternation becomes unproductive, the sounds are no longer distributed in a complementary way, but a tendency can still be discernible, and the method remains applicable. Sound changes can be reconstructed by IR as long as the number of "rule-confirming" forms outweighs the number of "exceptions." IR, however, cannot explain the causes of "exceptions" in most cases.

If a phonetic alternation occurs only irregularly, the reconstruction of the conditions could prove much more difficult. In most cases, a previous generalization of the rule can be assumed (i.e., the rule is recognizable in the forms with alternation once applied to the entire phonotactic system). King (1971:209) explained this procedure using the example of the Gothic *t-þ* alternation in participles already mentioned in Sect. 4.1.2.1 and thus arrived at a rule that is quite close to Grimm's law. However, as he (1971:211) himself admitted, this is only possible with some additional assumptions,[28] giving the impression of a purposeful reconstruction. A certain degree of regularity of an alternation is therefore the basic requirement for IR of sound change.

Determination of Conditions | The conditions of sound change are reflected in the phonotactic distribution, which necessitate a focus on the phonetic environments. Hoenigswald (1944) began to incorporate additional linguistic information (namely *phonemic distribution*) "with no morphology involved" (Hoenigswald 1944:79) into his approach. His approach was therefore also referred to as the "distributional approach" (Miranda 1975:303), but his method belongs to the morphological approach since the distributions merely extend the morphological method. His approach will be referred to as the MORPHOLOGICAL-DISTRIBUTIONAL APPROACH in this thesis. The

[28] King (1971:206 and 211) assumes, for example, another pronunciation of the Gothic letters ⟨*b,d,g*⟩ than the majority of Historical Linguists (cf. Braune and Heidermanns 2004:61-78).

approach starts with the consideration of an alternation's distributions, as seen in Fig. 4.7.

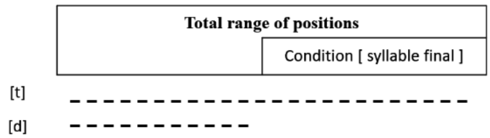

Figure 4.7: Distribution of the German final devoicing. Scheme according to Hoenigswald (1944:79).

Morphology is only involved here through morphophonemic consideration. The result is the automatic alternation /d/ [-A] ~ /t/ [A] for the phonemes, which is discernible for speakers (cf. Ringe 2003:246). From this observation, the hypothesis that "the phoneme of limited distribution was once free, and there has been conditioned sound change resulting in phonemic merger" is inferred (Hoenigswald 1944:80). It should be noted that by assuming a Ø-morpheme, which necessarily takes the position "total range of positions" (Hoenigswald 1944:81), one can also represent phonetic loss. As an example, Hoenigswald (1944:81) cited the English sound pair /k/ [#_n] ~ /Ø/ [#_n], which is found in the alternation *knowledge*:*acknowledge*.

Abstracting Single Sound Changes to Form Sound Laws | The simple application of the morphophonemic method to sound correspondences leads to single phonemic sound changes. However, since the actual condition of sound change is often sub-phonemic, an additional consideration of similar sound changes is necessary. The three single sound changes *b* > *p* \ _#, *d* > *t* \ _#, and *g* > *k* \ _# can thus be summed up to C[+plosive, +voiced] > C[+plosive, -voiced] / _#.

The prerequisite for such an abstraction is that the sound correspondences show a PARALLELISM with respect to their development or distribution of distinctive features; this development can then be distinguished from other sounds with the same distinctive features. The aforementioned sound change cannot be postulated if, in addition to the three voiced plosives, other voiced plosives exist that have not undergone this development. In practice, the internally identified sound changes, which actually belong to the same group of sound changes, often do not show exactly the same conditions, so the principle of parallelism may be concluded even if the conditions or even the sounds are merely similar. This can also lead to misconceptions, which can only be demonstrated through the Comparative Method (cf. Hockett 1958:469). In principle, abstracting can be carried out at the time of the determination of the sound pairs and distributions (see Lehmann 1962:103).

Reconstructability | The morphophonemic IR can be used to reconstruct conditioned merger or loss, as long as some morphology is involved, but it can seldom be used to reconstruct an unconditioned merger. A statement about sound change is implicitly derived from the basic hypothesis. In this way, an indicative determination of sound change can be made synchronically (Hoenigswald 1944:85–86 and 1946:142):

- automatic alternation in paradigms (→ merging or loss by conditioned sound change)
- regular, non-automatic alternation in paradigms (→ merging or loss by conditioned sound change with secondary sound change)
- free alternation in paradigms (→ phonemic split)
- free alternation in paradigm with Ø (→ older automatic alternation with loss)
- quasi-complementary distribution (→ borrowing and – if there is a "overlapping portion" – phonemic split)
- conspicuous distribution of paradigmatic and isolated wordform (→ loss of juncture)

Example Algorithm According to Marchand (1956) | The presented procedure of the morphophonemic approach can be specified differently in its details, for example in the determination of the sound-cha nge direction or the cognate identification. To summarize the method and to illustrate it with an example, the algorithm of Marchand (1956) will be presented here together with a concrete example. Marchand was one of the first to describe the formal principles for the application of IR in the case of sound changes. For this purpose, he (1956:246–250) defined five premises[29] and transformed them into a formal algorithm for the IR of *phonemic splits*:

[29] His premises include the basic hypothesis, predefining cognate pairs, determining sound correspondences, identifying the direction of the sound change, and determining of conditions.

1. A split is assumed if some formal conditions are met:
   a. The forms of the allomorphs must have the "same phonemes in as many positions as they have different ones" (1956:247), and these must be in the same order.
   b. The same alternation must occur in at least three morphemes or the morphophoneme in at least two morphemes.
2. Determine phonetic environments of the potential allophones.
3. Exclude all environments that occur with both sounds. A complementary distribution is obtained.
4. Apply the basic hypothesis: The phonemes in question have belonged to the same phoneme at an earlier stage. Other hypotheses may be possible.
5. If point 4 is impossible, only speculative hypotheses can be made. This applies, for example, to the case of merging with Ø.
6. The reconstructions are checked to see how many wordforms (especially anomalies) are explained by them.

As far as the last point is concerned, Marchand calls for a measure of "correctness." As such, either the methodological accuracy or the number and value of wordforms (especially anomalies) that the reconstruction can explain (Marchand 1956:246) is used. As an example of his method, he mentions the alternance of Lithuanian /s/ and /sʲ/, illustrated by the example of the paradigm of *liesas* 'thin' (Tab. 4.F).

Table 4.F. Partial paradigm of Lithuanian liesas 'thin'.

liesas 'thin'	**Singular**	**Plural**
Nominative	/lʲɛ**s-**as/	/lʲɛ**sʲ-**ì/
Dative	/lʲɛ**s-**am/	/lʲɛ**sʲ-**ɛmz/

The following premises are applied to this alternation:

1. Premises apply to this alternation.
2. /s/ appears only before back-tongue vowels and non-palatal consonants; /sʲ/ appears before front and back vowels and palatal consonants (Marchand 1956:249).
3. The condition "before back vowels" is excluded. A complementary distribution arises with /s/ before non-palatal consonants and /sʲ/ before front vowels and palatal consonants.

4. Assuming a loss of front vowels and palatal consonants after each /s^j/, which appeared before back vowels, the pre-Lithuanian allophony rule holds: /s/ is pronounced as [s] before back vowels and non-palatal consonants and as [s^j] before front vowels and palatals (Marchand 1956:250).
5. –
6. To review, the definite and indefinite wordforms of the paradigm of *geras* ‘good’ are compared:
   a. nom.pl.fem. /g^{j}ɛ̃ros/ with def. /g^{j}ɛ̃ros^{j}os/ < */g^{j}ɛ̃rosjos/
   b. acc.pl.fem. / g^{j}ɛràs / with def. /g^{j}ɛràs^{j}as/ < */g^{j}ɛràsjas/

   This reconstruction explains the palatalization of /s/ in the definite wordforms. These are based on the third-person pronouns (nom.pl.fem. /jõs/ and acc.pl.fem. /jás/)) and can therefore be considered as the most plausible reconstruction according to this approach.

Marchand’s method is in line with the IR scheme. Points 1 to 3 belong to the PRESUPPOSITIONAL PHASE, which extract the relevant data from the material. The following INDUCTIVE PHASE (pts. 4 and 5) is a build-up phase, which produces hypotheses, and a CONCLUSIVE REDUCTION PHASE (pt. 6) reduces the possible assumptions. Marchand strived to describe the procedure of IR as formally as possible. However, his conditions for determining suppletion were “rather arbitrary” (Miranda 1975:289), and Chafe (1959:478) indicated that he “accepts the hypothesis quoted without considering alternative possibilities.” In his example, Marchand omitted other possibilities of palatalization and left the notion of “anomalies […] that are explained by it” largely to the readers’ assessment. It should also be mentioned that point 3 presents a very simplified point of view concerning the data. Due to loanwords, neologisms, analogical levelling, and further sound changes, a covering phase arises in the linguistic history very soon after the incidence of the “wanted sound change” *s* > *s^j*. In this phase, it is possible that no “complementary conditions” remain after a strict sifting out of non-complementary environments. This point must work without a strict selection; otherwise, only phonological rules can indeed be reconstructed through this method. In fact, the examples listed by Marchand do not have a great time depth and even in cases such as Goth. ***triu*** (nom.) ~ *triwa* (dat.) ‘tree’ have a rather quasi-allophonic nature from a synchronic point of view.

4.2.2 Distributional Internal Reconstruction

4.2.2.1 Phonotactics

Since the establishment of the morphological-distributional method, the question of whether a purely distributional method (i.e., not based on alternations in allomorphs but on phonotactics) exists has been raised repeatedly (Hoenigswald 1944:81, cf. Miranda 1975:303). In other words, is the information that *g* does not occur in word-final position sufficient to reconstruct a sound change *g* > *k* / _#? This question is particularly relevant for weakly inflectional or non-inflectional languages. In these languages, a morpheme-final *g* could, by chance, never occur in word-final position (e.g., due to missing "suffix-less" word stems).

Instead, most linguists seem to support the opinion that where no alternation ensues, a "reconstruction is not possible" (Fox 1995:159–160). Clearly, this applies to wordform reconstruction. According to Miranda (1975:303), final devoicing can also be reconstructed purely phonotactically "with as much certainty as when alternation occur." However, he did not take into account that such a combination (e.g., only [k] and no [g] in final position) could also have arisen in other ways (see Tab. 4.G). A loss of final *g* is another explanation for this distribution.

Table 4.G. Fictive distribution of *g* and *k* before (1) and after (2) the sound change *g* > *Ø* / _#. The final distribution (2) corresponds to the scheme of final devoicing (see Tab. 4.C).

(1)	_V	_#
g	+	+
k	+	+

g > *Ø* / _#
→

(2)	_V	_#
g	+	-
k	+	+

In the morphophonemic IR, these cases would be intercepted by morphological alternations. For Hoenigswald (1944:81), the only trace left by sound changes beyond morphology would be "a peculiar gap in the list of phoneme clusters" (see also Jeffers and Lehiste 1979:47). As an example, Hoenigswald cites the English phoneme /h/, which has been restricted in its distribution by loss in several positions (e.g., in pre-consonantal onset positions as in Old English *hnecca* 'neck' and *hring* 'ring'). The problem with this approach is that a large number of such "gaps" exist in each language, the majority of which cannot even be attributed to sound change. Hoenigswald would have had to use this method to reconstruct not only an *h* > Ø / #_C but also *l*, *r*, *m*, *n*, or k^w > Ø at this position. One might object here that it is possible, in principle, that an *#ln-* may have actually existed in a very old phase of “Pre-English” and that this phonetic sequence did not re-emerge in the period after the sound change. Arguments that attribute each absence “more likely” to sound change than to coincidence are used for more or less justified reasons (cf. Freeze 1976:200). Nevertheless, such a development would by no means be compelling. If new phonemes arise (e.g., *k* > *c* / _i) the occurrence of the phoneme may be restricted depending on the condition of the corresponding sound change (in the example, *c* is missing in final position). A reconstruction could lead to the assumption of *c* > Ø / _#. A better situation is available with more conclusive data when there is a complementarily distributed assimilation such as a palatal before front vowel and velar before back vowel as in the example in Hock (1986:537–538). The status of phonetic assimilation here is already implied by the condition of potential sound change, thus eliminating many alternatives.

4.2.2.2 Sound Frequency

With the help of the distributional method, the possibility of an internal reconstruction of an unconditional merger is occasionally discussed (Hoenigswald 1944:81). However, it is not the phonotactic that plays a role here but rather the relative sound frequency (cf. Jeffers and Lehiste 1979:47). In the merging of sounds, the frequency of the resulting sound increases significantly. For instance, different ancient Greek front vowels merged with *i*, making it the most frequent vowel in modern Greek. Furthermore, after the pre-Sanskrit merger of the old mid vowels and the syllabic nasals with *a*, the phoneme *a* occurred in Sanskrit texts with a frequency of 19.78%, which is twice the frequency of the second most frequent sound (Hoenigswald 1965:99). This frequency can hardly be explained synchronically, other than through a conditional or unconditional merger. However, the sounds before the sound change cannot be determined so easily. Low-frequency sounds, which are primarily handled as lost sounds, may also have been newly formed by loanwords or other sound changes and may have a low frequency only by chance.

Internal reconstructions based on frequencies of sounds or sound sequences have been proposed repeatedly but involve morphology to a greater or lesser extent (e.g., Adelaar 1983 with a tendency to *pattern reconstruction*). Nevertheless, the phonotactic approach is often criticized — not without good reason — because in strict implementation, there is no procedure but only an assumption that the frequencies were once more balanced (Hock 2013:15).

4.2.2.3 Symmetrical Phoneme Systems

A similar approach based on the symmetry of phoneme systems has been discussed. Fox (1995:163) assumed phoneme systems of parallel structures (e.g., an equal number of front and back vowels or of short and long vowels). As an example, he cited the merger of Proto-Indo-European **a* and **o* to Germanic **a*, which created a gap at the position of **o*. This gap indicates an unconditional merging. Fox (1995:163–164) argued that these positions are often reanalyzed to restore symmetry. In the Germanic example, **a* could be rated as the back posited equivalent of **e* (Fig. 4.8).

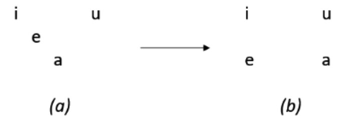

Figure 4.8: On the left, the phonetic positions of the Germanic vowels, and, on the right, the systematic re-analysis according to Fox (1995:164).

In contrast to the morphophonemic approach, distributional IR is not able to reconstruct wordforms but only sound change. Especially on the zenith of generative grammar, there was occasional talk about the reconstruction of phonological rules instead of sound changes (e.g., Freeze 1976:201). Thus, in the case of a distributionally reconstructed final devoicing, it is no longer possible to say which wordform can be traced back to a voiceless or a voiced final consonant. Without any alternating forms or morphological information, there is no possibility of reconstructing the pre-form of German word *und* [ʊnt] 'and'.

Another problem with this method is the lack of a *direct* relation between phonotactics and sound change. The sounds [b] and [d] do not occur after [s] in very many languages except due to a sound change of *b* > *p* or *d* > *t* / *s*_ (cf. Hoenigswald 1944:80).

4.2.3 Pattern Reconstruction

The term *pattern reconstruction* is used in this thesis to refer to all types of internal reconstruction that are designed for a reconstruction of root structures or at least take into account any semantic aspect for a phonotactic unit. Instead of "cognate morphs," this method compares "structural patterns": that is to say, structural patterns such as *CV̄* and *CVC* are traced back to a common pre-linguistic pattern **CVC* (Latta 1978:3–4). In the literature, this method is referred to as the "structural method" (Hjelmslev), "pattern reconstruction" (Latta 1978:3), or "distributional method" (Borgström 1954:276). It is often traced back to Saussure with his theory of *coefficients sonantiques* or to Streitberg (1896:81–82), who explains the development of Proto-Indo-European long vowels by means of a Pre-Proto-Indo-European compensatory lengthening (e.g., Borgström 1954:275). The exact definition of "pattern" varies. Ringe (2003:257) defined it as a "pattern of alternation that perform the same grammatical function" and cited the Germanic *ablaut* as examples. In the same way as the distributional method of Sect. 4.2.2, pattern reconstruction works with phonotactics; however, this is not used for reconstructing sounds but rather phonotactic structures with some (partly only theoretical) morphological value.

For pattern reconstructions, a BASIC HYPOTHESIS adapted to "patterns" applies, which Ringe (2003:258) referred to as the "assumption that the paradigms in question must have been morphophonological parallel." Two or more patterns, such as the participials' root structures of the Germanic *ablaut* series one to three, trace back to an original pattern (Tab. 4.H). From this reconstruction, a sound change for the third *ablaut* series N̥ > *u*N can be inferred. Since the validity of pattern pairing generally is more speculative than the pairing types in Sect. 4.2.11, the basic hypothesis ultimately has a weaker explanatory value. The possibility of identifying cognates is mostly restricted to their similar (grammatical) function.

Two approaches to pattern reconstruction are presented below: the "distributional" and "conclusive method" according to Borgström (1954), as well as pattern reconstruction according to Latta (1978), which was presented in detail in his doctoral thesis.

Table 4.H. Illustration of pattern reconstruction using the Germanic ablaut patterns. The Gothic participles refer to the Pre-Gothic pattern *CVRC* : *CØR̥C* (with C = consonant and R = sonorant).

	***Ablaut* Pattern**	**Goth. Infinitive**	**Goth. Participle**
I	CViC : CiC	b*e*i*t-an*	b*i*t*-an-s*
II	CVuC : CuC	b*i*u*g-an*	b*u*g*-an-s*
III	CVNC : CuNC	b*i*n*d-an*	b*u*n*d-an-s*
	↓		
	*C*V*RC : *CØR̥C		

4.2.3.1 Pattern Reconstruction According to Borgström (1954)

Borgström (1954:276) mentioned two special cases of IR called distributional and conclusive method. Both approaches can be attributed to pattern reconstruction. The first approach is based on the distribution of entities, just as the phonotactic method is. Borgström defined “distributional method” as follows

> *A certain pattern, e.g. a rule for the distribution of entities, which is not motivated in the given system, may be motivated, though not rendered simpler in itself, under certain systematic conditions which may then be assumed for an earlier stage of the language.*[30]

He did not give an example to substantiate his claim but demonstrated this method with his postulated “Rule for the Distribution of Vowels” (RDV) in Proto-Indo-European. The Proto-Indo-European *schwebeablaut* and zero-grade show indications of a Pre-Proto-Indo-European syncope, which led to reconstructions such as Pre-Proto-Indo-European *CVCVC developing into CCVC or CVCC, depending on the stress position. Borgström (1954:279) argued that in this way, the distribution of consonants in Proto-Indo-European was not arbitrary, but not every consonant could take any position. *i̯ could only occur in vowel environment and a purely plosive root structure *ptk* was not possible. Borgström (1954:282) concluded that the position of the vowel depended on the number of consonants:

- CC > CVCV
- CCC > CVCCV
- CCCC > CCVCCV
- CCCCC > (C)CVCCVCCV

By attaching suffixes, the position of the vowel changed (cf. 1954:282) and the present suffix -i was added in later times:

- *$h_1$äs-tä ‘(it) is’ > *$h_1$est > Proto-Indo-European *$h_1$ésti, but
- *$h_1$sä-ntä ‘(they) are’ > *$h_1$se-nt > Proto-Indo-European *$h_1$sé-nti.

According to Borgström (1954:278–279), the cause of the *schwebeablaut* is therefore morphological and not due to a preceding syncope. Nevertheless, Borgström

[30] Borgström (1954:276)

(1954:286–287) himself had to admit that a large number or words violate his rule, which he can explain only “by analogical levelling and by the new zero-grade. Several morphological and derivational forms must be later than the RDV.” As long as the assumption of more recent developments is not confirmed, the rule remains hypothetical. A feature of this method is that it does not use the phonotactics of synchronic languages to reconstruct a sound change *a* > *b* but to reconstruct the phonotactics of the pre-language.

As a second special case of IR, Borgström mentioned the so-called “conclusive method”. This case applies if one draws “further conclusions about the structure of the earlier system” after having reconstructed parts of the older language (Borgström 1954:276).The method he described is a regularizing method, which is not attributed to IR in this thesis. Under the term “regularizing method,” I understand the attempt to eliminate irregularities in a *reconstructed language*. The basis of this method is the assumption of a corrective rule that must be applied to reconstructed wordforms. If the motivation for the rule is a typological assumption, it is also called the “intergenetic method” (Greenberg 1978:84). Reconstructions based on reconstructions are not to be discarded from a methodological view; it should be noted that each further layer of reconstruction depended on the probability of the correct reconstruction of the underlying layer. If, in theory, one was able to reconstruct the entire vocabulary of a language — which is rather unlikely given the possibility of word loss and neologisms — one can determine phonological rules to some extent in purely descriptive terms. Regularizing methods are associated with a higher error rate if they require a revision of the empirical data. So far, in comparative and internal reconstruction, we have induced a theory (e.g., a phonetic law) on the basis of empirical data (mostly historically attested cognates), which, in turn, served to reconstruct the proto-language (see Sect. 2.1.1). In regularizing reconstruction, the empirical material is revised by changing reconstructions (e.g., by replacing the reconstructed series Proto-Indo-European **d* - **t* - **d^h^* by **d* - **t* - **t^h^* with purely typological rules). In these cases, there is indeed a danger of circular reasoning. As examples, Borgström (1954:278) gave three premises from Proto-Indo-European derived from externally (and internally) reconstructed wordforms:

1. The Proto-Indo-European vowels *i* and *u* (as well as *m̥, n̥, r̥, l̥*) are interchangeable with *y, w* (*m, n, r ,l*) and originally seem to have only had consonantal functions;

2. Proto-Indo-European *ē,ō,ā* as well as *a* and *o* which does not interchange with *e* are believed to have developed from *e* in the neighborhood of laryngeals;
3. The remaining Proto-Indo-European vowels *e,o* (and lengthened *ē,ō*) are interchangeable and seem to have developed from the same *e* as under (2).

From these premises, Borgström concluded a monovocalism in Pre-Proto-Indo-European, which he rendered with **ä*. The three premises were inferred inductively, which is, in principle, legitimate for reconstructed languages although a reconstruction of the complete vocabulary (i.e., empirical data) is unlikely and the inductive theory (e.g., each *ā* < eh_2) cannot be verified. This approach only becomes problematic when the empirical material itself is adapted to the theory. For instance, contrary to premise 1, **w* also occurs in non-vocalic environment (e.g., **u̯reng-* 'turn, twist' LIV 639). Additionally, a laryngeal environment cannot be derived from the cognates not for every *a*, *ā*, *o*, and *ō*. In fact, according to premise 1, the wordforms Skt. yájate 'sacrifice' and Greek *ἅζομαι hadzomai* 'revere' lead to Proto-Indo-European **Hi̯aǵ* (LIV 200), which violates premise 2; premise 2 leads to **Hi̯h₂eǵ-*, which violates premise 1 (Fig. 4.9; cf. Tichy 2009:38). Premise 3 is based on the hypothesis of etymological allomorphy. However, the Proto-Indo-European *ablaut* is synchronically a morphological alternation, and Proto-Indo-European does not give insight into the condition of the possible sound change. Consequently, there may have been several vowels, which then collapsed to *e* and *o* and were morphologically redistributed. Finally, it should be noted that the impression of monovocalism may be due to the simplified approximation of proto-languages (see Sect. 2.1.3).

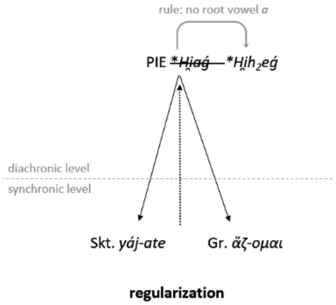

Figure 4.9: Illustration of regularizing methods using the cognate pair Skt. *yájate* 'sacrifice' and Greek *ἅζομαι hadzomai* 'revere'. The reconstructed word root **Hi̯aǵ-* is adapted (i.e., regularized) to the assumed rule that there was no root vowel a in Proto-Indo-European.

Borgström's premises are ultimately based on regularizing approaches and, thus, do not enable the conclusion of a monovocalism. Assuming the correctness of premises 1 and 2, the conclusion is based solely on premise 3 and works here methodically according to the morphophonemic method with a reconstructed dataset. A special case of IR is ultimately not present here.

4.2.3.2 Pattern Reconstruction According to Latta (1978)

Latta (1978) described the method of *pattern reconstruction* formally for the first time in his doctoral thesis but attributed it to Saussure and Benveniste (cf. Latta 1978:3). In his thesis, he spoke of "structural" or "relational patterns," by which he (1978:93) meant "any of various formulae by which a language marks grammatical distinctions." *Relation patterns* are more abstract than morphemes, but according to Latta (1978:119), they are language-systematic real and "rooted in empirical observations of the way languages change."

Method | The thesis of Benveniste (1935:149–151) started with the already mentioned *schwebeablaut*, an observation that there are pairs of words with the same or similar meaning, only differing by the position of the root vowel before or after the resonant: **u̯erg* : **u̯reg*, **teru̯* : **treu*, or **genH* : **gneH*. Latta (1978:79) began the first step with a DEFINITION OF A PATTERN *TeRT* : *TReT* (where *T* is a consonant and *R* is a sonorant). The initial difficulty of pattern reconstruction lies in the recognition of so-called "allopatterns" (Latta 1978:88) like *TeRT* : *TReT*. In the simplest case, functional or semantic similarity or equality supports pairing cognates and allopatterns that can be identified in this way. In the case of Saussure's theory, one had two relational patterns in late Proto-Indo-European:

1. *V̄* (*full grade*) and **ə* (*zero-grade*, here as *schwa indogermanicum*)
2. *eR* (*full grade* with R = resonant) and R (*zero-grade*).

The first point is traced back to the older patterns **eH* > V̄ and *H* > **ə*, making it possible to connect both points (Latta 1978:96). The systematic pairing of the two patterns is denoted by two colons (::); thus, in Latta's (1978:97–98) notation, they are represented as follows: *eR::R* and *V̄::ə*. There is usually a functional difference between the two structures of a pattern (so-called CANONICAL SHAPES; i.e., *eR* and *R*). In the case of Saussure, these are the functional differences between the two *ablaut* stages. In Benveniste's theory, the meaning of the difference between *TeRT* and *TReT* is unknown. In the work of Latta (1978:96–97), an inflectional paradigm is understood as "a bundle of relational patterns" of root and "desinence" (e.g., *base* + *nom*, where "base" stands for all possible structural forms; that is, canonical shapes of roots). Allomorphy only affects the specific instances of the pattern, not the general character of the pattern (cf. Latta 1978:96).

The second step of pattern reconstruction is to interpret the data to form a hypothesis explaining structural anomalies. In the simplest case, as in Saussure (1879), one canonical shape can be interpreted as a morphophonological variant of the other. However, this is not always possible if there have been reinterpretations, analogical levelling, or changes to both canonical shapes (see examples in Latta 1978:104–108). In the case of Benveniste, the forms are interpreted as "terms of ablaut" (Latta 1978:82). Benveniste suggested starting from a root *TER* and a suffix *-ET*, with the forms of *TeRT* showing a loss of the suffix's vowel and the forms of *TReT* showing a loss of the root's vowel. He regarded this rule, rather, as a synchronic rule without necessarily assuming an original ***TeR-eT*. The postulation of an abstract system of roots and suffixes, however, implies a diachronic aspect. Latta (1978:82) traces the pattern pair *TeRT::TReT* to an older pattern ***TeR+et*, just as allomorphs are traced to an older pre-morpheme in paradigmatic IR.

The third step is "to explicate the nature of the linguistic changes that the pattern has suffered" (Latta 1978:116), which includes the listing of sound changes. This brought Latta (1978:116) to the following algorithm for pattern reconstruction:

- Establish the existence of systematic pairing of morphemes of different canonical structure.
- Infer the original shape of the pattern on the basis of (a) a direct comparison of the pairings (in the case of pattern split) or (b) the careful consideration of the available structural evidence.
- Provide a detailed account of the linguistic changes that have affected the original pattern and resulted in the pattern change.

Reconstructability | In parallel with the sound-change typology of Hoenigswald (Sect. 4.1), Latta (1978:98–100) distinguished between two types of reconstructability of the so-called *relation pattern change*:

- PATTERN SPLIT:
  A pattern of the proto-language is decomposed into two new patterns of the younger language stage. This corresponds to the examples of Benveniste and Saussure.

- PATTERN REANALYSIS:
  This case occurs when the pairing of the canonical shapes is formally changed, or the pattern is lost (Latta 1978:99). Changes in pairing usually occur by reinterpretation. For instance, the locative of the Proto-Indo-European verbal noun results from the suffixation of -*i* to the verbal root + derivative suffix: *base+es::base+es+i*. In Latin, the vowel *e* was interpreted as a theme vowel of the thematic inflection (**douk+e+t* > *dūcit* '[he] leads') and *s* (> Lat. *r*) was attributed to the final suffix: *base+e::base+e+si* (Jeffers 1977:19–20). If the pattern is lost (pattern loss), the canonical shapes are no longer interpreted as related. This case occurs in lexicalizations.

Unlike a sound change, the reinterpretation of pattern change cancels out the continuity between the original and the new pattern. The new pattern **V̄::ə* does not continue *eR::R* but replaces it (cf. Latta 1978:101–102). Reconstruction of *pattern reanalysis* is rarely possible because internally, there is often no evidence of pattern change. On the other hand, a pattern loss can be reconstructed similarly to a pattern split. The difficulty lies in showing the systematic relationship of the old patterns.

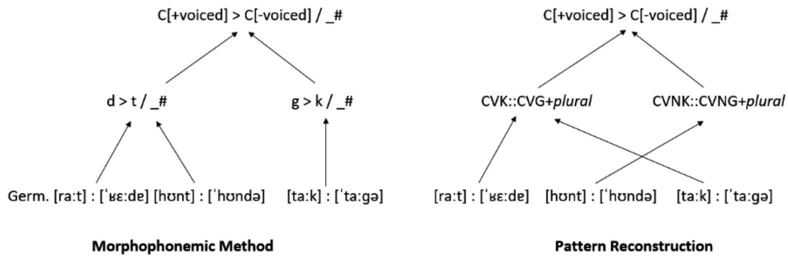

Figure 4.10: Comparison of morphophonemic IR and pattern reconstruction exemplified through final devoicing in German words (*Rat* 'wheel', *Hund* 'dog', and *Tag* 'day'). Doing pattern reconstruction, abstracting the sounds in question is done earlier.

Review | Pattern reconstruction and structural IR are widely accepted as separate forms of IR although IR and the structural method are equated by Anttila (1968:166); otherwise, the pattern reconstruction is rated as "an extension of the standard methodology of Internal Reconstruction" (Fox 1995:178). Ringe (2003:257–259) placed them on equal footing with the morphophonemic method and saw in it a similar degree of uncertainty for the latter. Latta (1978:4) himself concluded that "the method must be considered to be 'safe.'" I see the major methodological difference between pattern and IR in the order in which they conduct abstraction. The morphophonemic IR first compares allomorphs intra-paradigmatically, then compares the sounds and sound environments with other morphemes inter-paradigmatically, and finally abstracts the respective sound changes. Pattern reconstruction first abstracts the morpheme to pattern and then begins the comparison (cp. Fig. 4.10).

In Saussure's example, the morphophonemic method also leads to the correct sound changes. A problem with pattern reconstruction, however, is the historical dissimilarity of sound change and pattern change. Since pattern changes can only be regarded as reinterpretations, no sound change can be directly derived from them, such as *eH:H* > $\bar{V}$ > *ə*. These must be inferred from the corresponding morphemes (as in Latta 1978:110). This raises the question of whether abstracting to patterns is useful for reconstructing sound change at all. In addition, only relatively few sound changes, such as syncope, can result in a pattern split at all (cf. also Latta 1978:116). Sound changes that do not affect patterns can be sorted out through early abstracting. I observe advantages in early abstracting in more complex morpheme structures. Internal reconstructions of full reduplications can lead to erroneous results if strictly using the morphological-distributional method (see Fig. 4.11).

Pattern reconstruction can also refer to old morphological formation patterns if a set of semantically or functionally similar words has a phonetic commonality that distinguishes them from others (see Sect. 3.6.3). In this case, however, a reconstruction of sound change is not possible.

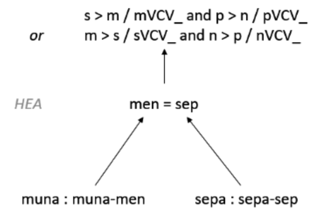

Figure 4.11: Illustration of the morphophonemic IR applied for fictive reduplicated wordforms. The reduplicated morphemes *men* and *sep* are incorrectly regarded as allomorphs of the same pre-morphemes.

I do not consider a strict separation of the two methods (as proposed by Latta, for example) as mandatory. Furthermore, the basic hypothesis of both methods (one with allomorphs, one with allopatterns) can be regarded as variants. The similarity of both methods can be seen if we take a look at the definition of pattern as "any of the various formulae by which a language marks grammatical distinctions" (Latta

1978:93). With the inclusion of “grammatical distinctions,” which is normally expressed morphologically, we also work inherently with allomorphs and move only slightly away from the original definition of the basic hypothesis.

4.2.3.3 Pattern Reconstruction According to Austerlitz (1986)

Austerlitz (1986:189) tried to “contribute to refining and to redefining the [internal, A.B.] method itself” and identified three issues to define this method (1986:184):

- Faulty distribution within the paradigm: “the distribution of any feature which is statistically frequent enough to pass for normal and to be rare enough, at the same time, so as to stand out slightly”;
- A statistical judgment on textual or lexical frequency;
- A judgment based on the “*Wörter-und-Sachen*” match.

Austerlitz (1986:185–189) demonstrated this method using examples of Gilyak, an isolated language in Eastern Siberia. For this purpose, he constructed special paradigms consisting of word families with common lexical (‘*Wörter-und-Sachen*’) and phonetic features, such as the paradigm **dV-* ‘arm or part thereof’ in Tab. 4.I. In the first step, it was assumed that these words were originally polymorphemic and contain a morpheme **dV* ‘arm or part of thereof’. Austerlitz then went further and analyzed other sound sequences. The palatal nasal *ñ* of the words (5–7) was compared with other words containing *ñ*. All of them could be subsumed under the meaning ‘sharp, pointed’, whereas the *m*-containing words meant ‘characterized by (motion of) a joint or joints’ (Austerlitz 1986:187). For rare phonetic sequences in a language’s phonotactics, he (1978:3) hypothesized an original morpheme boundary. The rare sound combination *st* within Finnish roots was therefore derived from an older morpheme boundary **s-t* (e.g., Finnish *osta* ‘buy’ < **os-ta* containing *osa* ‘part’).

Table 4.I. The so-called ‘paradigm’ for Gilyak **dV* ‘arm or part of thereof’ according to Austerlitz (1986:186).

	Word Form	**Meaning**
1	*damk*	‘hand’
2	*dot*	‘arm’
3	*domχ*	‘elbow’
4	*dořpŋ*	‘forearm’
5	*dakñ*	‘fingernail’
6	*doqñ*	‘claw’
7	*duñmŋ*	‘finger’

Although Austerlitz (1978:4 and 1986:188–189) himself pointed out the difficulties and ambivalences in determining meaning, he did not seem to view the method as problematic despite its subjectivity and susceptibility to error, which resulted from the loose semantic and phonetic similarity of those words. Especially in the last reconstructions in his 1986 paper with its abstract interpretations, the similarity to the decompositional method (see Sect. 5.1) became unintentionally obvious. Connections such as Finn. *sydän* 'heart' and *syy* 'fibre' or *syyä* 'deep' (1978:4) do not have any semantic proximity. This may also be the reason why his approach was hardly applied in practice in later IR papers although it was described as the "most valuable methodological account" (Isebaert 1991:219).

4.2.3 Onomatopoetic Internal Reconstruction

As a form of "reconstruction interne," Naert (1957:7) described two onomatopoeic methods which should complement Bonfante's typology (Sect. 2.3.1). Although they are performed without linguistic comparison, they subsequently attracted little attention as IR methods. To Campbell and Grondona (2007:24), these methods have "very little connection with the method of internal reconstruction as recognized today," leading them to omit a closer treatment.

4.2.3.1 Methods

Naert called the first approach METHODE DES ONOMATOPEES QUI N'EN SONT PLUS. It was assumed for a word that there was an original onomatopoeia as a proto-form phonetically transformed by sound change. For example, Naert (1957:6) cited Swedish *gök* [jøːk] 'cuckoo'. Due to the tendency of many languages to name birds after their birdcalls, he set an onomatopoetic "archetype" in accordance with the basic hypothesis, without naming it specifically. Since this is a neologism in the pre-language, there is indeed a proto-form in the proper sense. For the next step, Naert argued as follows:

> *Ceci nous amènerait à restituer une occlusive vélaire derrière le [j-] actuel (pour la détermination de la nature de cette occlusive, sourde ou sonore, etc., il faudrait évidemment avoir recours à d'autres méthodes). On serait aussi en droit de très fortement soupçonner que le [ø:] non plus n'est pas primitif,*

mais que le vocalisme ancien a été plus « sombre », plus « caverneux ».[31]

Similarly, he (1957:7) derives the English noun *pipe* from *piːp and *gape* from *ga:p.

The second method is called LA MÉTHODE DES ONOMATOPÉES QUI ONT RÉSISTE (ibid.) and can be used when two onomatopoetic words that refer to the same *signifié*. In Ainu, there are two words for "to bark": *mik* and *mek*. According to Naert (ibid.), *mek* is the better onomatopoeia and therefore probably the older form, which suggests a sound change *e* > *i*.

4.2.3.2 Review

Naert's onomatopoetic methods are not easy to apply. First, the potential vocabulary is limited to a few semantic fields (animal names, verbs of sounds, etc.). The basic hypothesis is therefore only valid for a limited scope.[32] Second, the original "proto-form" is only very vaguely defined. The inter-linguistic variation of onomatopoeias with the same meaning is by no means to be underestimated. The method assumes that people reproduce a certain sound in a "universally" uniform way and that differences are caused by sound change. This hypothesis is difficult to support empirically. In addition, sound changes are often irregular in onomatopoeias. Thus, an identified sound change does not have to represent a diachronic sound law.

The lack of a uniform "archetype" presents a major issue for automating the onomatopoetic method. One possible way is to define an "average form" of the word in question. For instance, the proto-form of *cuckoo* is determined as the "average form" of the word in all languages. The differences between this "average proto-form" and the word in question are possible sound changes. Finding an appropriate average form is an unsolved issue.

[31] Naert (1957:6). Translation: "This would lead us to restore a velar occlusive behind the present [j-] (for the determination of the nature of this occlusive, dull or sonorous, etc., one would obviously have to resort to other methods). One would also be likely to strongly suspect that the [ø:] is not original either but that the old vowel was "darker," more "cavernous."

[32] According to Grammont (1901), the number of original onomatopoeias is much larger. However, these would be unusable for this method as long as there is no way to determine the original sound.

4.3 Types of Grammatical Internal Reconstruction

In addition to the reconstruction of phonology, morphology, and lexicon, comparative reconstruction also includes the reconstruction of semantics, syntax, and grammatical inflectional categories. Morphology and lexicon are partly included in wordform reconstruction, which is based on the reconstruction procedure of sound change. In addition, they contain some semantic or grammatical components, which, just as semantics, syntax, and inflectional categories, must be reconstructed in a different way. These types are subsumed in this thesis under the term "grammatical Internal Reconstruction." Morphological wordform reconstruction may give hints to the existence of grammatical categories in a proto-language. From forming an equation Greek perfect = Old Germanic preterit = Skt. perfect, one can conclude a separate Proto-Indo-European category "*perfect*." The question of the function of this category can only be done through inferring methods. If, for example, language A has the word order *S-O-V*, language B uses both *S-O-V* and *S-V-O*, and language C has free word order, individual phenomena in the individual languages may indicate that there was an innovation in language B, and the original word order may have been *S-O-V*. However, this conclusion cannot be inferred from the equation itself (cf. Fox 1995:191–192). Thus, it is the reconstruction of syntax based on alternations that remain questionable. The syntactic alternation "*He gives the book to her*" and "*He gives her the book*" offers few criteria to reduce both variants to one pre-form. In this context, it seems possible to assume that prepositional phrases are younger and thus the first variant has to be the younger one. However, from an internal point of view, this assumption is not otherwise supported. The issue of making appropriate assumptions has led to a negative attitude toward syntactic IR in the literature (e.g., Lightfoot 1979:161).

Those attempting to reconstruct grammatical subsystems in a purely internal way are forced to find comparable material (i.e., correspondences) within the language (e.g., in word order or idioms deviating from the language's norm). Since these subsystems are far more open to transformation and analogy than phonology and morphology, conclusions about them are much more difficult to reach. It is therefore hardly surprising that the subject scope of IR is not infrequently restricted to the field of phonetics and morphophonemics (cf. Birnbaum 1970:98) and that the method is sometimes regarded as "invaluable in historical morphology" (Shevelov 1964:5).

4.3.1 Archaistic Internal Reconstruction

An intuitive way to project a relative chronology within a language's system is the diachronic projection of linguistic elements that are synchronically perceived as "archaisms" (see Chafe's stylistic variants in Sect. 4.2.1.1). Bonfante and Naert were among the first to describe this special form of IR, which Bonfante called "method of the 'fase sparita'" and Naert called "méthode des archaïsmes conventionnels."

4.3.1.1 Methods

According to Bonfante (1945:135–136), for example, a word is considered the older of two synonymous dialectal terms if it is the "dying word" (*fase moribonda*). As examples, he cited the French *clore* and *fermer* 'to close', the Italian *capo* and *testa* 'head', and German *Kopf* and *Haupt* 'head'. Nevertheless, without an insight in the historical texts, the determination of the "dying word" is hardly possible. The frequency of the words in question serves as a means of identifying the "dying word," but this does not allow any conclusion about older language stages. For instance, the German word *Sofa* 'sofa, couch', which emerged in the 17th century (KLUGE 855) is more common today than *Faulbett*, which appeared for the first time in the 16th century (KLUGE 280). However, in modern German, *Sofa* is in more frequent than *Kanapee*, which emerged as a vogue word in the 18th century (DUDEN 413).

What can be grasped "internally," however, is the subjective synchronic sense of archaism, which corresponds to Chafe's (archaic) stylistic variants. In addition, there was an attempt to extend this rule to syntactic synonymy, where falling productivity refers to the older construction (see Birnbaum 1970:93). Again, the question of synchronic measurement of diachronic changes in productivity remained unanswered.

Thus, instead of the "frequency" of words, "style" (i.e., connotation) served as the determinant for Naert (1957:6–7) and subsequent linguists. That meant that the "conversationally" connoted word can be interpreted as the pre-form for the younger. As an example, he cited Swedish *timme, timma* 'hour' ~ *urtīma* 'which falls outside regular time; extraordinary (about a session)'. From this word pair, it could be inferred that the original meaning was 'time' and not 'hour', as well as a sound change *īm* > *imm* (Naert 1957:6). The example is rather atypical, however, since ideally both lexemes are synonymous and differ only in their archaic connotation.

Milewski (1973:102), who uses the term "method of exceptional forms" for this method, gave the alternative Polish forms *w niebiosach* ~ *w niebiesiech* 'in the heavens' and *w Prusach* ~ *w Prusiech* 'in Prussia' as examples. The forms ending in *-ach* are originally borrowed from the feminine declension and today constitute the regular ending. The forms ending in *-ech* are obsolete and historically older.

4.3.1.2 Review

One problem with the archaistic method is that anything unproductive (or, in the case of a proto-language, thought to be so) can be treated as "archaistic" or residual, and, thus, a circular conclusion may arise. To avoid such a circular argument, Bauer (2009:26, 29) proposed a multi-step procedure specifically for proto-languages:

- reconstruction of the main characteristics of the proto-language
- identification of "deviations"
- evaluation of deviations in terms of linguistic differences and parallels
- identification of residues versus innovations
- evaluation of the archaism of residues
- determining what other constructions or forms represent the same linguistic horizon

Bauer (2009:25) thus concludes that Proto-Indo-European *mihi est*-constructions (possessive dative with copula) and impersonal verb constructions originate from to the same Pre-Proto-Indo-European syntactic construction. The core of the method can be found in the first step, the "main characteristics," which, in the case of Proto-Indo-European, is a characteristic of a nominative language. If structures deviating from the norm show similarities, these are assigned to the pre-language. In this way, the validity of the result depends on the correct determination of the main characteristics and the common features.

Naert (1957:6) excluded poetic archaisms in his definition since these are more likely to belong to the "chronology of texts" (and thus to another reconstruction method according to Bonfante's typology). This objection is quite justified since these "literary" archaisms are deliberate references to historical texts (e.g., poetry or religious writings). This brings the synchronic aspect of IR into question. On the other hand, this separation from "conventional" and "literary" archaisms cannot be clearly demarcated, nor is it clearly differentiated by Naert.

In this thesis, the method is attributed to grammatical IR although Naert reconstructed sound changes in his examples as well. In fact, Naert reconstructed his examples by applying the morphophonemic method with semantic pairings: He declared *timma* and *-tīma* to be allomorphs (cp. also Chafe's special forms of stylistic variants). Other linguists, such as Birnbaum (1977:10–11), have only seen this method as a means of reconstructing syntactic structures. However, the question of the extent to which the connotations perceived synchronically as "archaic" in fact represent an older stage remains. It is possible, in principle, for a marked ***and*** an unmarked construction to be assigned to the pre-language (see Lightfoot 1979:158). In this method, synchronic subjectivity seems to be interpreted as diachronic reality. Hock's (2013:12) criticism that the results can hardly be called reconstructs is justified since the procedure is ultimately only a selection from two alternatives and not a trace back to a preform.

4.3.2 Suppletive Internal Reconstruction

Through synchronic suppletions, morphemes can be reconstructed from pre-languages since suppletion is the result of a morphological merger (Fox 1995:186–188). On a lexical level, the pre-morphemes **go* and **wend* can be inferred from the English suppletion *go* ~ *went*, while at grammatical level, the influence of analogical levelling complicates this process (Fox 1995:187). The meaning of these morphemes can only be specified very fuzzily with **wend* 'go *or the like*', but this is also often true for the Comparative Method. In contrast, I recognize a difficulty in the recognition of suppletion, as, for example, the Latin paradigm *fero* 'I bring, carry' ~ *tulī* 'I brought, carried' ~ *latus* 'brought, carried (participle)' would thus lead to three morphemes. In fact, however, *tulī* and *latus* (< **tlāt-o-)* are etymologically related (Meiser 2010:108).

4.3.3 Grammatical Internal Reconstruction According to Kurylowicz

Kurylowicz (1964 and 1973) was one of the first linguists who devoted the most extensive attention to IR methods in more complex subsystems. His works focused on changes in functional usage and opposition. At a morpho-phonological level, such a change can be paraphrased as follows (cf. Kurylowicz 1964:14):

Given a phoneme Φ_1 (negative-neutral or unmarked, e.g., PIE **e*), a phoneme Φ_2 (positive and marked, e.g., PIE **ē*), and a phoneme Φ (neutralized, e.g., PIE **ə*), which occur in morphological alternance: in the base forms or simplices, the vowel is Φ_1, and in the derivatives, it is either Φ_1 or Φ. Now, the process of *polarization* (i.e., the tendency to obtain the strongest possible contrast between derivatives and base words) takes place. The neutral Φ (*ə*) is now interpreted as the neutralization of Φ_2 (*ē*), which provides the basis for polarization, and in the derivatives, Φ_1 can be replaced by Φ_2, while Φ_1 remains confined to the base forms (see Kurylowicz 1964:14 for more details).

According to Kurylowicz (ibid.), the neutralized form allows a reinterpretation of the morphological procedure and the determination of the *condition of neutralization* forms the basis of reconstruction in this method. For example, to determine the origin of the Proto-Indo-European *ablaut*, it is therefore necessary to start from the origin of the zero-grade (i.e., the neutralized form). The "original point" lies here in the accent positions of *e* and *Ø*. In this way, Kurylowicz (1964:15–16) concluded that the derivation sequence $T_1eiT_2 > T_1iT_2 > T_1oiT_2$ was the Pre-Indo-European development.

This concept can then be transferred to other grammatical categories. Kurylowicz's structuralist approach became even clearer when he (1964:16) wrote that one "of the most important tasks in reconstruction is to establish the position of a form within the system, the pertinent morphological opposition in which it participates." The function of a linguistic unit — a word, morpheme, or other unit — usually has a context-independent, *primary function* (e.g., the plural morpheme with the primary function "plural") and a context-dependent, *secondary function* (e.g., the collective meaning of a plural morpheme), which can only be inferred from the context (cf. ibid.). In his example, the category "plural" would be *inflexional* (i.e., part of paradigms), while the category "collective" is *derivational* for certain nouns. A derivational category can become inflexional (*grammaticalization*), or an inflectional category can become derivational (*lexicalization*). Based on this, Kurylowicz derives the following rule, whose function is to be compared with that of a basic hypothesis:

> *As a rule it is legitimate to consider inflexional forms as former derivatives. As regards the rules of derivation they may represent an extension of older, narrower rules (in which case we are concerned with grammaticalization) or, on the contrary, they may be the result of the lexicalization of broader rules (even of inflexional categories). [original emphasis, A.B.]*[33]

[33] Kurylowicz (1964:17)

Assume a grammatical form F with the primary function Φ_1 and the secondary function Φ_2. The formal renewal F > F' may produce a split in which the primary function receives the new (i.e., productive) form (F' = Φ_1), and the secondary function may continue to be used with the old form (F = Φ_2); however, in this usage, it may also be slowly displaced by F' (cf. Kurylowicz 1964:26). In some cases, the old forms have become lexicalized. For instance, the Latin neuter plural ending *-a* with the primary function "plural" was replaced in Italian by the masculine form *-i* but remained productive in its secondary function as collective nouns: Italian *muro* 'wall' (< Latin *murus* masc.) ~ *muri* 'walls' (< Latin *muri* masc.) and *mura* '(city) walls' (cf. Kurylowicz 1973:85).

4.3.3.1 Relative Chronology

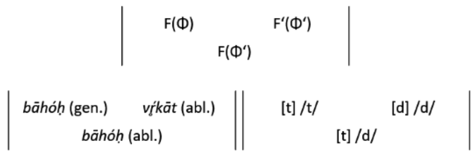

Figure 4.12: Representation of syncretism according to Kurylowicz (1964:27) with F and F' for word-forms and Φ and Φ' for functions (upper display) and examples of syncretism (lower displays): on the left, case syncretism in Sanskrit *bāhú-* 'arm', while *vŕ̥ka-* 'wulf' formally still differ between ablative and genitive (Kurylowicz 1964:27); on the right, final devoicing as an example of phonetic syncretism.

It should be noted that only a relative chronology is reconstructed in the examples mentioned above. Relative chronology is, to Kurylowicz (1964:10), the "chief aim" of IR. For such a chronology of grammatical categories, primary and secondary functions are decisive factors (Anttila 1973:344). In the example of Italian *mura*, a phase 1 is set with only one plural *mura* (F), a phase 2 with the emergence of the form *muri* (F') from other paradigms, and a phase 3 in which the form F' became productive (i.e., more used) and pushed back form F to the secondary function of being a collective noun. The possibility of a relative chronology arises from the panchronic and panlingual character of the hierarchy of primary and secondary (i.e., derivational) functions (Kurylowicz 1964:26). It is important to note that the concept of relative chronology is merely a relative chronology "of attested morphological devices (morphs)" and not a chronology of functions according to Kurylowicz (1964:28). Consequently, a secondary development such as *masculine* > *feminine* does not give any idea of the origin of the category feminine, because the older, unattested feminine nouns could have been displaced by new derivational forms. The actual task here would be to determine which function of a morpheme is the primary function, which can only be done by determining the opposition within the system, not within a particular context (cf. Kurylowicz 1964:26–27). In the case of syncretism — by which he understands, among other concepts, final devoicing — Kurylowicz proposes the representation displayed in Fig. 4.12.

The formally and functionally differentiated forms F(Φ) and F'(Φ') have primary functions. The mixed form F(Φ') indicates a secondary function. At a phonetic level, a final devoiced sound is one such mixed form, whose form F corresponds to the phone and the function Φ' to the phoneme (see lower panels in Fig. 4.12). On a morphological level, this can be illustrated by case syncretism. In some Sanskrit noun classes, the genitive form displaced the ablative form and took over its functions, while in other stem classes, the separation was formally preserved.

The suffix *-āt* in *vŕ̥kāt* 'wulf (ablative)' confirms the independent morphological status of the ablative, so the ablative function can be determined as a secondary function of both functions of the suffix *-oḥ*.[34] In reconstruction, however, one must expect secondary developments that may entail "phenomena of differentiation" (i.e., any changes in the formal-functional relationship; Kurylowicz 1964:29). This includes the substitution of forms with secondary function for forms with an identical primary function. Kurylowicz (ibid.) saw the use of universals as "chief keys" here.

4.3.3.2 Reconstruction

An example of the reconstruction of a grammatical system is cited in Kurylowicz (1964:22–24) for the reconstruction of the functions of the Proto-Indo-European aorist. Formally, the Greek and Sanskrit wordforms of present, imperfect, and aorist are inherited from Proto-Indo-European, but they differ in their functions. Ancient Greek is an aspectual language, in which the present and imperfect tenses render the imperfective aspect and the aorist tense expresses the perfective aspect. Sanskrit has a temporal system in which the present and imperfect forms express simultaneity (with the narrated moment = narrative) and the aorist tense an anteriority.

Applying Kurylowicz's concept of opposition Φ_1:Φ_2 and neutralization, the mapping in Tab. 4.J holds for Greek.

[34] In accordance with Kurylowicz's notation (1964:27), a *sandhi* variant of the suffix is used.

Table 4.J. Greek aspect system according to Kurylowicz's concept of 1964. The opposition imperfective : perfective aspect is formally expressed by imperfect and aorist.

imperfective	perfective	neutralized
Φ_1	Φ_2	Φ
imperfect	aorist	present
ἔ-λειπ-ον	ἔ-λιπ-ον	λείπ-ω

Table 4.K. The temporal system of Sanskrit according to Kurylowicz's concept (1964). The opposition of simultaneity (=non-anteriority) and anteriority is formally expressed by present and aorist.

simultaneity	anteriority	neutralized
Φ_1	Φ_2	Φ
present	aorist	imperfect
sárp-at-i	*á-sr̥p-at*	*á-sarp-at*

Since aspect does not matter in the present tense, it is functionally "neutralized" (Φ). In contrast, a language with a temporal system, such as Sanskrit, "neutralizes" the narrative tense since there is no anteriority (cp. Tab. 4.K). In Greek, the opposition forms *ἔ-λειπ-ον e-leip-on : ἔ-λιπ-ον e-lip-on* show only a "*minimal* difference" (Kurylowicz 1964:23). In contrast, the Sanskrit opposition pair *sárp-at-i : á-sr̥p-at* shows a strong formal discrepancy and "is therefore deprived of its autonomous distinctive value" (ibid.). The formal similarity of the aorist form to the imperfect form in Sanskrit indicates a more recent development, so it can internally be assumed to have developed from a secondary function. If in Proto-Indo-European, the aorist had the same primary function as in Sanskrit, "we should expect its stem to admit primary as well as secondary endings" (1964:24).

Kurylowicz's IR is concerned with finding discrepancies between the functional and structural levels. On a structural level (i.e., on the *signifiant* side), the same morphological structure applied to both languages (Fig. 4.13 lower display). Formally, the positivity is marked with *ablaut* (*p*). In the present form, the "past markers" (*n*+*o*; i.e., augment and secondary ending) are neutralized. On a functional level, the structures in the two languages look different (in Fig. 4.13 with *x* = perfective, *y* = preterital, *z* = anterior). The triangle in Sanskrit is "rotated" at the functional level compared to the structural level. This discrepancy is the starting point of Kurylowicz’s reconstruction. Semantically redundant morphemes (i.e., *morphonemes*[35]) are particularly suitable for finding such discrepancies (cf. Kurylowicz 1973:88–89). From the opposition *positive : negative* emerges the morphological structure *zero : n*. In the Greek aspect system, this corresponds to the structure *zero : zero-grade*; the suffix *-i* is, here, “*redundant* in relation to other tenses” (Kurylowicz 1973:89 provides another example).

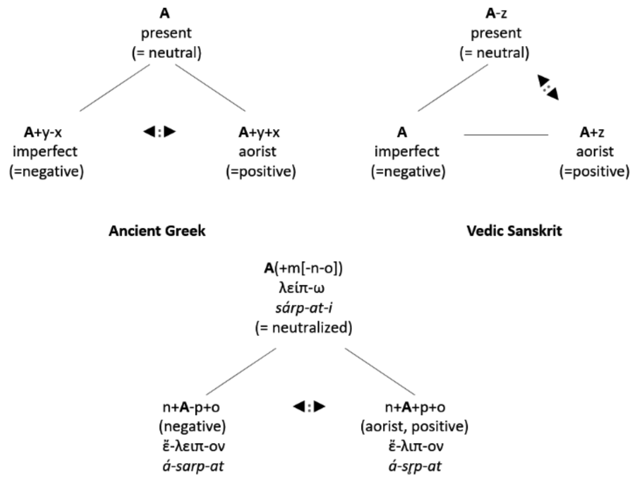

Figure 4.13: Functional structures of temporal oppositions in Greek and Sanskrit with x for *perfective*, y for *preterit*, and z for *anterior* (upper displays) and the formal structure of time oppositions in Sanskrit and Greek with *A* as root, *n* as augment, *m* as primary ending, *o* as secondary ending, and *p* as ablaut (lower display).

4.3.3.3 Review

Kurylowicz’s chosen example of grammatical IR shows a simplification of often more complex structures. Especially the reconstruction of the Proto-Indo-European root presents and root aorists may be much more difficult (see also Winter’s statement in Kurylowicz 1964:35). Kurylowicz’s reconstruction model has also been criticized. By comparing the Greek and Sanskrit tense systems, Boretzky (1975:50) observed a massive violation of Kurylowicz’s own IR definition, as Kurylowicz used comparative approaches. It seems justified to ask whether, in reconstructing the Pre-Sanskrit model, he was not already influenced by the well-known Indo-European result.

[35] see also Anttila (1973:342): “empty of meaning […] is thus a morphoneme […] independent meaning is thus a morpheme. […] in internal reconstruction it is important to distinguish between semantically empty morphs from morphs with semantic or syntactic function.”

Anttila (1973:342) saw a problem with grammatical IR in the fact that it requires the complete grammar of a language with all its details, which inevitably leads to problems in practice. He himself only ventured to make statements about relative chronologies in grammatical IR. To him (ibid.), issues of productivity and correlation with basic vocabulary play a crucial role here. In Finnish, for example, an inflectional type that does not formally distinguish between noun and verb can be determined (e.g., *tuule-* 'wind' and 'blow'). This type is unproductive and belongs to basic vocabulary. Consequently, it is regarded as older than the suffixal inflectional types. In English, in contrast, formal differentiations between noun and verb are rather unproductive: *fill* to *full* or *redden* to *red*. From this, one would suspect "that Finnish has grown flexion and English has lost it" (Anttila 1973:342).

4.3.4 Givón's Principle

Givón formulated an aphorism, later termed "Givón's principle" (Fox 1995:193), for syntactic IR: "yesterday's syntax is today's morphology" (Givón 1971:413). The early modern English term *methinks* is derived from a univerbated Anglo-Saxon phrase *me þynceþ* (Wischer 2011:362) and indicates an impersonal usage of the verb *think* in Pre-English. Givón's principle, however, remains uncertain. As Lightfoot (1979:160) suggested, although object-verb composites such as *nutcracker* correctly refer to a Pre-English word order O-V, they are still productive today, and verb-object composites such as *breakwater* have also existed in English for centuries (for a detailed discussion, see Jeffers and Lehiste 1979:119–124, Comrie 1980 and Dezső 1980).

4.3.5 Typological Internal Reconstruction

4.3.5.1 Typological Methods of Bonfante (1945)

Attempts to reconstruct sound laws with the help of linguistic universals or "usual" sound changes were already found in Bonfante (1945:144), who mentioned two possible methods. The first one he called the METHOD OF THE USUAL PHONEMIC CHANGE or "l'analisi 'interna'" (Bàrtoli 1935:417). Through the juxtaposition of Latin *senex* and Greek *ἕνος hénos* 'old', he identified a Proto-Indo-European wordform starting with **s*- since the sound change *s* > *h* is "the usual phonemic change" (Bonfante 1945:85) compared to the other direction. This assumption is based on studies by contemporary linguists such as Jespersen (1922:263) and Schwyzer and Debrunner (1939:56), but the universality of the direction *s* > *h* has been repeatedly questioned. The development of *h* to a sibilant can be attested in some Japanese dialects (cf.

Naert 1957:2). In addition, the method of the usual phonemic change is actually not a real reconstruction method but rather an inferring method and tends to be discussed as such (e.g., Fox 1995:195–196). The method of usual phonemic change serves to reduce the reconstruction possibilities, whether they have been reconstructed comparatively or internally. Without already identified cognates, its application is not possible.

Bonfante differentiated between this typological method and the more abstract METHOD OF GENERAL LINGUISTIC EVOLUTION. The latter method includes "universal" linguistic changes, such as the loss of a dual, the formation of articles, or the evolution from aspect to tense (cf. Bonfante 1945:85). Jakobson (1941:76) coined the term "allgemeine Lautgesetze" (*general sound laws*) for such universal and panchronic changes and went as far as to see those as valid even for glottogonic assumptions. The examples described by Bonfante are in fact developments that are now rather interpreted as parallel developments of a European *sprachbund*, the so-called *Standard Average European* (Haspelmath 2001; before him, among others, Shimomiya 1974). Bonfante (1945:86) also referenced the "general tendency of language to shorten words." He cited Sturtevant's (1917:173–175) statistical survey, which counted the syllables in the Gospel of Matthew in ancient and modern languages. The ancient Greek and Latin text have 39,000 and 37,000 syllables respectively, while the German (34,000), French (33,000), or English (29,000) Gospel have a significantly lower amount. In my mind, there is a distorted picture here. On the one hand, most Proto-Indo-European wordforms consist of fewer syllables than the modern Indo-European languages, which is, among other things, due to laryngeal vocalization. For instance, the PIE word for 'star', **hstḗr*, in Latin became *stella* and *ἀστήρ astḗr* in Ancient Greek, which became Spanish *estrella* and Greek *ἀστέρας astéras* respectively. The number of syllables increased with each stage. According to the "general tendency," we would expect more syllables in older stages.

On the other hand, Bonfante's example, Goth. *habaidedeima* compared to English *we had*, is not very meaningful since Germanic weak verbs are a historically young development, arising from the synthesis with an auxiliary verb (*-*ded*- 'did'; cf. Bech 1963:5). Thus, the phase of syllable reduction was preceded by a phase of syllable addition. The same is true for Bonfante's (1945:132) statement about the "general tendency to go from the concrete [semantics, A.B.] to the abstract, not vice versa (or very rarely so)." Evidence of semantic development *abstract* > *concrete* is indeed far more frequent than "rarely," as can be seen in the Anglo-Saxon *hund* 'dog' to modern English *hound* 'hunting dog'.

Bonfante himself speaks only of "tendencies" that retain their validity despite counterexamples. However, their status as "universal" remains completely unclear, especially since the given examples can rather be attributed to the convergent development of Standard Average European. There is also no reason to assume that languages in general do not build up an aspect system or a dual form or they tend to dismantle them as soon as one has emerged. Notwithstanding, it remains an open question whether this method is a reconstruction method or an inferring method. As an inferring method, its use would be comprehensible. If one assumes the rule *aspect > tense*, then one can conclude an aspect system in the proto-language from the existence of both systems in the daughter languages. However, the method is not suitable as a reconstruction method for each tense language, a language stage with an aspect system cannot be presupposed.

4.3.5.2 Typological Method of Kurylowicz (1964)

As a reconstruction method, universal sound laws would have to postulate the same pre-sound for each occurrence of a particular sound or sound combination. This assumption is difficult to prove empirically but has nevertheless been adopted by different linguists. Even IR skeptic Miranda (1975:304) considered some universal sound laws to be legitimate, such as the back-projection of word-final nasal vowels to a combination of vowel and nasal (*VN#* or *NV# > Ṽ#*). The universal character of this sound change was generally supported by later studies (Ferguson 1966:59 and cf. Sherzer 1972), but it was not explicitly classified "unrestricted universal" (cf. Greenberg 1978:63).

Approaches to establish this "universalist" method as part of IR can first be found in Kurylowicz (1964) although he does not distinguish between reconstruction method and inferring method.[36] Later authors such as Givón (1999a:120–121) considered these kinds of methods "analogical-abductive reasoning" and thus inferring methods. Since Kurylowicz primarily focused on grammatical reconstructions of IR, this method is subsumed under this category, although it can be adaptively applied to phonetic reconstruction. A more phonetically oriented model was presented by Greenberg (1978), but his methods were not designed exclusively for IR.

Determining a hierarchy of semantic or syntactic functions of a morpheme forms the basis of Kurylowicz's method. Kurylowicz (1964:29) saw the UNIVERSAL LAWS (also "panchronic laws") as possible "chief keys to reconstruction," which are

[36] In fact, the morphophonemic approach and other methods are based on observations from the way how language change proceeds generally and universally. From this point of view, it is sometimes assumed that IR is generally based on language typology (e.g., Joseph 2010:53).

mostly based on a *hic-nunc-ego* situation. He therefore dedicated a large part of his papers on IR to determining or deriving concrete diachronic universals (see 1973:70–86). As examples of universal laws, he (1964:29–30) stated:

- *iterative* > *durative* (present) > (general or indetermined) *present*
- *static verb* > *perfect* > *indetermined past* > *narrative tense*
- *collective* > *plural*[37]

If a collective noun already functions as a plural under defined conditions, this can only be interpreted as its secondary semantic function, which can be expressed, for example, by morpho-syntactic incongruity (Kurylowicz 1973:73). Alternatively, if there are parallels between iterative and durative forms, the durative forms can be derived from the iterative forms but not vice versa.

While Kurylowicz listed exclusively linear universals in his 1964 paper, he seems to assume rather circular universals in his 1973 paper. From concrete historical developments, such as the emergence of adjectives in Lithuanian, he (1973:82–84) induced, for instance, a "syntactical circulation between the noun and the adjective" and between verbs and adjectives (see Fig. 4.14). However, he does not elaborate on how these circulations can be specifically useful for IR. According to these circulations, an attributive adjective can be traced back either to a verb or to a noun. Tracing this developmental path for reconstruction no longer seems possible. The generality of these circulations is also unclear for languages that do not have a noun-adjective distinction.

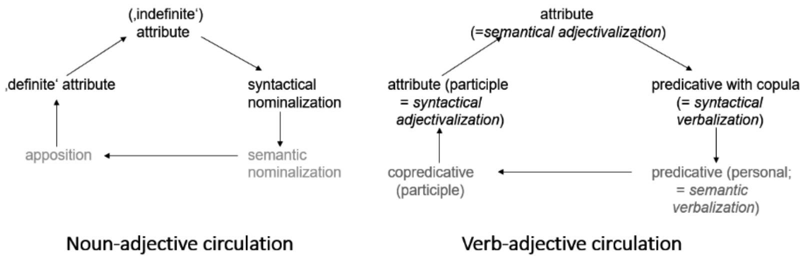

Figure 4.14: Kurylowicz's circulation of nouns and adjectives (1973:82) and verbs and adjectives (1973:83-84). Adjectives are colored in black, nouns in grey and verbs in dark grey. Own depiction.

[37] For Kurylowicz (1973:73), this "seems to be a diachronic universal."

4.3.5.3 Typological Method of Greenberg (1978)

Greenberg's method works with a *state-process model* (1978:67–85) in that two states A and B can merge either into one another, or unilaterally, or not at all. For the development of nasal vowels from nasals, the states in Fig. 4.15 apply. Greenberg (1978:75–76) incorporated *implicational universals* into his graphic, which means that some sound changes imply some preceding sound changes universally. For instance, nasalization of high vowels implies — in accordance with empirical data — preceding nasalization of low vowels but not vice versa. He (1978:75–85) also discussed other methods of integrating typological knowledge into the state-process model to resolve the problem of too few universals. In this context, he suggested adding empirical probabilities to this model to build a kind of *Markov chain model* in which each identified sound change was assigned a likelihood. On the whole, such an approach is difficult to implement due to the large number of possible transitions.

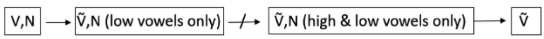

Figure 4.15: Transition of vowel-nasal combinations to nasal vowels according to the state-process model of Greenberg (1978:74).

Another approach proposed was by the method of "dynamization of subtypology," which only considered languages that were typologically similar (e.g., languages with voiceless vowels). According to a synchronic implicational universal, each voiceless low vowel implies the existence of voiceless high vowels. This suggests that in languages containing voiced and voiceless vowels, either all vowels became voiceless at the same time under a certain condition or the high vowels existed first, followed by low vowels (Greenberg 1978:78–79).

4.3.5.4 Review

Kurylowicz (cf. 1964:30–31) saw the discovery or recognition of linguistic-historical universals through empirical studies as a decisive factor for the future of historical linguistics and IR. Contemporary linguists were already speaking out against this optimism. Lehmann (in Kurylowicz 1964:33) objected to *semantic* universals, arguing that they had "not found general statements decisive in attempting to account for phonological change" (see also Hamp's stance on the suitability of universals for historical research in Kurylowicz 1964:36). Anttila (1973:350) warned that "flimsily established universals are taken at face value, and strong inductive claims are made without much empirical foundation." Fox (1995:198) recognized his *panchronic laws* as being of interest for IR as an exclusionary inferring method "though not much use can be made of them in this form."

To me, universals generally seem to be inconclusive for concrete internal reconstructions since in IR, after all, only a relative chronology "of attested morphological devices" and not of functions is established (Kurylowicz 1964:28). The universal *collective* > *plural*, for example, does not state that the Proto-Indo-European plural form **-es* goes back to a collective **-es* since the plural category may have existed before the emergence of the morpheme **-es*. The universal merely asserts that "at some point" the plural function originated from a collective function. A concrete plural form may also have a different origin. It may have developed from a dual form, suffixation, or univerbation. Thus, the universal is, on the one hand, practically impossible to verify and, on the other hand, irrelevant for reconstruction purposes because it does not lead to a concrete reconstruction. Even if the hypothesis is correct, it says nothing about the original plural form or collective form, except that it functioned as a collective. Thus, this indicates that we have not gone beyond the basic hypothesis.

The lack of a strict separation between reconstruction methods and inferring methods also led to a split in the usage of the term "universal" in the literature. For example, Jeffers (1976:8) called the inclusion of universals in the Comparative Method a "new form of internal reconstruction" even though his examples are clearly described as comparative.

Similarly, Greenberg's methods seem to have found little acceptance in reconstructive linguistics. The few phonetic *(near-)universals* barely offer potential for wordform reconstructions, whereas a probabilistic model provides too many potential pre-forms for one wordform. His theory was examined several times by subsequent linguists (e.g., Croft et al. 2011) but was never used as a means of IR. The necessary demand for *unidirectional* and *irreversible universals* remained; otherwise, "they lose much of their predictive value" (Fox 1995:205–206).

4.3.5 Semantic Merger According to Hoenigswald (1965)

Internal methods for reconstructing semantics have been proposed in the literature as well. A phonetic alternation is missing in these cases, so those methods do not deliver any statements about the sound change. An overview of semantic methods of IR is therefore dispensed in this work, and the procedure utilized is only exemplified by Hoenigswald's (1965) approach. If only the wordforms (i.e., *signifiant*) merge, homonymy arises, which — if the semantic ranges are far apart — may suggest a historical sound change or a borrowing:

> *Homonymy through borrowing or sound change, on the other hand, is semantically only an accident, so that in most cases the „filling-in“ would have to range over the entire table of meanings (to fill in the gap between mead ‚a beverage‘ and mead ‚meadow‘, one would have to combine distributions which are not typically combined in the distribution of other nouns; hence one would conclude, even if no other information were available, that mead and mead are homonyms from sound change) [original emphasis, A.B.]*[38]

If Hoenigswald's hypothesis about homonyms is correct, however, this would mean that among the polysemous lexemes of a language individual sounds appear more frequently in comparison to the rest of the language's vocabulary. This could be interpreted as an indication of a merger and thus of a diachronic sound change. Corresponding research has not yet been conducted thus far.

Nevertheless, Hock (2013:11–12) saw here a methodological similarity to traditional IR, but he required more additional assumptions. As IR tries to trace two expressions (*signifiant*) with the same content (*signifié*) back to an original pre-form, Hoenigswald's approach tries to unify two contents by using the same expression. Ultimately, however, a different assessment of English *mead* 'meadow' : *mead* 'flesh' and Russian *mir* 'peace' : *mir* 'world' leads to the result that the first homonymy is considered less likely (Hock 2013:12).

[38] Hoenigswald (1965:69)

5. Scientific-Historical Aspects

In the last chapters, the different definitions and methods of IR as they are used today were presented and analyzed. In addition, further methods were brought forward in the course of the history of science, which can be formally treated under the term “Internal Reconstruction,” or deviations in the definition or procedure may have been found. An adequate presentation of IR is only made possible taking into account its scientific-historical aspects. This chapter is intended to shed light on and analyze the changing views IR itself.

5.1 Prelinguistic Internal Reconstruction Exemplified by Cratylus

In the period before the 19th century, “etymology” represented the branch of science that sought to obtain the “true” (Greek *ἔτυμος étymos*) meaning of a word. Its approach was based on the view that the original (i.e., “true” form of a word) had decayed through the course of time. The methodical attempt to recover this original form and meaning can be regarded as a precursor of IR, unless other languages were included.

Within this thesis, the methodology of the prelinguistic “old-etymologists” can be exemplified through the dialogue *Cratylus* of Plato (~425–347 or 348 BC). The extent to which the work is suitable as an example of pre-scientific IR may be questioned. In the literature about *Cratylus*, it is discussed which etymologies are attributed scientific value[39] and which must be regarded as polemically rejected ideas (e.g., Heitsch 1984:36–43 and Rehn 1982:23–34); scholars also question whether *Cratylus* is to be regarded as a linguistic work at all or rather primarily as an epistemological work (cf. Derbolav 1972:50–52). In the dialogue, Socrates indicates that the used methods can lead to one deceiving oneself (CRATYLUS 428d) and in the final discussion, the question of “correct” etymology is critically narrowed down and taken back to the more essential question of epistemological correctness (Derbolav 1972:40). Whether Plato actually advocated the etymological method outlined

[39] The main representatives of this view are Sedley (2003a:28–50) and Hiller (2001:35 and 40–41).

in *Cratylus* or merely reflected critically-satirically on a contemporary model to demonstrate the futility of etymological word interpretations is therefore left open here. Despite these criticisms, *Cratylus* represented a starting point for later etymologists, who expanded upon Plato's methods regardless of his actual intentions. Unlike later etymologists, who often held the view that all contemporary languages descended from Hebrew, Greek, or the so-called "Adamitic language" of the tower of Babel (cf. Haßler and Neis 2009:535–537), Plato did not assume an external proto-language and consequently did not deal with comparative methods. In considering the historical aspect of IR, *Cratylus* is relevant because it represents a kind of decompositional IR, which also characterizes the Austerlitz's method (see Sect. 4.2.3.3).

The work *Cratylus* was written in the form of a fictitious dialogue between the philosophers Socrates, Hermogenes, and Cratylus. The discussion starts with the question of whether the forms of words are arbitrary (i.e., based on agreement and habit) or whether they should be understood as set "imitations" of Platonic *ideas*. In the dialogue, Hermogenes represents the so-called "conventional" view and Cratylus the "naturalistic" one. Socrates, who leads the discussion, takes a mediating position and comes to the conclusion that the knowledge of truth cannot be achieved by means of words (CRATYLUS 435d–436b).

A central role in the Cratylian discussion is played by the so-called rule-setters (*νομοθέτης nomothétēs*), by whom one understands in the broadest sense all people forming new words and in a narrower sense those who formed the first words of a language (for more details, see Sedley 2003b). The rule-setters are compared with craftsmen who try to reproduce an archetype (a "Platonic idea," e.g., the abstraction of a hammer) with their product (i.e., the concrete hammer). The naturalists, on the other hand, assume that the rule-setters tried to depict the essence of the word's matter.

5.1.1 Method of Associative Decomposition

Word associations form the basis of many lay attempts at synchronic reconstruction. Behind this etymologized approach is the idea that semantically or phonetically more complex words are to be regarded as composites. Derbolav designated this procedure to the "metaphorical" level of language and describes it as follows:

> *… mit der Senkung des logischen Bewußtseinsspiegels öffnet sich der Wortauslegung ein lockerer Hof semantischer Assoziationen, die von der jeweils antizipierten Bedeutung des etymologisierten Wortes her ausgelesen und dann auf phonetische Ähnlichkeit mit seiner Lautgestalt hin durchkomponiert werden.*[40]

In this way, the word *ψυχή psychḗ* 'soul' is identified as that "what holds and guides the one's nature" and thus the word can be seen as a univerbation of the phrase *τὴν φύσιν ἔχειν καὶ ὀχεῖν tēn phýsin échein kai ochéin* 'to hold and guide nature'. This resulted in a pronunciation *φυσέχη psychéchē*, which was then pronounced as *ψυχή psychḗ* for euphonic reasons (CRATYLUS 400a–b). Plato's etymologies, constructed in such a way, are indistinguishable in their validity from other competing etymologies because of the arbitrariness of phonetic substitutions and deletions. This leads to three different etymologies for words such as *Poseidon* (CRATYLUS 402e and 403e), without one being clearly preferred. This makes it all the more clear that Plato understands the etymological attempt as being "correct" that has the greatest etymological information content (i.e., that describes their intended as accurately as possible; cf. Derbolav 1972:69). In this respect, Plato's procedure differs from other associative decomposing methods in which, for the most part, an etymology is considered authentic if the source and target words are as phonetically similar as possible.

According to Socrates, the reason for the phonetic discrepancy between the source and target words is, on the one hand, because of the rule-setters, who changed the original phonetic structure for euphonic reasons or for concealment (CRATYLUS 395e and 399). On the other hand, time is also a factor that ensures a perpetual flow of change (CRATYLUS 411c).

The whole method is also criticized by Socrates himself, who suggests that if one can put letters into words and take them out at will, every interpretation seems to be possible (CRATYLUS 414e); thus, the original intention of the rule-setters may

[40] Derbolav (1972:67). Translation: "… with the lowering of the logical level of consciousness, this is an unconstrained place for semantic associations that could be opened up to word interpretations, which are read out from the respective anticipated meaning of the etymologized word and then composed based on phonetic similarity to its phonetic form."

have been made completely unrecognizable. To curb the resulting arbitrariness and subjectivity of the results, he refers to the need for "measure and equity" in any etymological attempt (cf. CRATYLUS 418b). The method of reshaping the phonetic structure of a word in order to subsequently decompose it into other words also entails a certain *anachronistic paradox*. One looks for synchronic base wordforms in reconstructions projected back in time. These base words obviously have not undergone any phonetic transformations. If one assumes any past phonetic substitutions, one must also assume some transformations for the base words, which, in turn, makes it impossible to verify the correctness of the etymologies.

5.1.2 Method of Phonetic Physiognomy

The result of the associative decomposition is a group of basic words that cannot be further decomposed. The procedure has thus reached its limits. None the less, an attempt is made to continue this procedure by giving the remaining units (i.e., the sounds) a semantic concept. Through the semantics of letters and sound sequences, words were created that should "imitate" the referring things (cf. CRATYLUS 425). For instance, the sound *r* with its vibrant pronunciation expresses 'movement', which can be demonstrated by words such as *ῥεῖν rhéin* 'to stream' and *τρόμος trómos* 'tremble' (CRATYLUS 426c–e).

Phonetic physiognomy actually gleans its semantics from the meaning of the example words, making the method inherently circular. Robinson (1956:325) drew a comparison with the English word *ugly*, which may sound "ugly" to many speakers, but this view is based less on the phonetic structure of the word than on its meaning. The circularity of the method becomes apparent through phonetically similar words with opposed meanings. Phonetic physiognomy thus inevitably leads to contradictory results. The method can only be maintained if the material is adapted to the phonetic physiognomic result and one assumes a historical sound change for contradictory words (as done in CRATYLUS 434). All things considered, the phonetic physiognomic approach methodologically fails because the reconstruction that it brings about only serves to maintain its basic hypothesis.

5.1.3 Review

Since Plato did not assume linguistic relationships in *Cratylus*, none of its presented reconstruction methods can be attributed to the Comparative Method. From synchronic linguistic diversity within the speaker's community, a diachronic linguistic divergence is inferred (i.e., a sound change). From synchronic linguistic similarity between speaker's communities, diachronic linguistic convergence is inferred. This kind of reconstruction can be classified as External Reconstruction.

The method of associative decomposition can be called "Internal Reconstruction" in that it works purely synchronically without comparison with other languages. It can be compared with the semantic approach of the morphological method, but it shows a variety of differences. Whereas the semantic approach works mainly with (unproductive) derivations, the Cratylic approach starts from composites, which often violate the word-formation rules of their own language. For example, the words δύο *dýo* 'two' and ἀγωγή *agōgḗ* 'induction' cannot be formed into *δυογόν dyogón* (> *ζυγόν zygón* 'yoke', cf. CRATYLUS 418), neither derivationally nor by univerbation. Compounds are much less suitable for IR. Sound changes in compounds are far more often reanalyzed and regularized (cp. e.g., Middle High German *schuo**ch**m**e**chære* 'shoemaker' > *Schu**h**m**a**cher* by analogy with *Schu**h*** 'shoe' and *m**a**chen* 'to make'). If at least one component stem of the compound is no longer recognized, there is also the possibility of misanalysis (i.e., *folk etymologies*). The Cratylic etymologies fail methodically because they do not succeed in fully accounting for the phenomenon of sound change. The arbitrariness of the applied sound changes not only made each etymology unfalsifiable but also led to a temporal paradox since the synchronic words of the component stems are used as older proto-forms for the compounds. If, however, one assumed that the component stems had also undergone various, arbitrary sound changes, the concept of associative decomposition underlying this method would be deprived of the basis of reconstruction since the arbitrary sound changes could give rise to any phonetic form. The component stems would have had a different form at the time of the word-setting.

5.2 Linguistic Internal Reconstruction in the 19th Century

5.2.1 Pre-Neogrammarian Internal Reconstruction: Sustaining Old Approaches

Approaches comparable to those of *Cratylus* can also be found in the early phases of historical linguistics. Schlegel (1808:49) described a unidirectional *Stufengang* ("gradual progression") of languages from isolating via agglutinative to inflectional. This idea is also reflected in the works of Bopp, the first professor of Indo-European studies. Bopp tried, for example, to explain PIE **s* containing in future and aorist forms to be the verb **as* 'to be'. In doing so, Bopp — just as the protagonists of

Cratylus did — finally took the view that a proto-language consists of monosyllable roots and a process of “composition” has taken place in its daughter languages (cf. Delbrück 1880:57).

Table 5.A. Verbal paradigm of ancient Greek *εἶναι éinai* ‘to be’ with different wordforms of the 1st pl. in Attic, Doric, and Ionic dialects.

	Singular	**Plural**
1.ps.	εἰμί *eimí*	ἐσμέν *esmén* (Attic), ἐμέν *emén* (Doric), εἰμέν *eimén* (Ionic)
2.ps.	εἶ *éi*	ἐστέ *esté*
3.ps.	ἐστί(ν) *estí(n)*	εἰσί(ν) *eisí(n)*

In addition to the method of associative decomposition, approaches similar to the MORPHOPHONEMIC METHOD can be found before the 19th century. The Byzantine grammarian Georgios Choiroboskos from the early 9th century may have been the earliest scholar to have used such methods (cf. Curtius 1873:146), treating the irregular ancient Greek paradigm for the word ‘to be’ (Tab. 5.A). Choiroboskos tried to explain the juxtaposition of the forms of the first-person plural by means of a historical development:

καὶ ἀπὸ τοῦ εἰμὶ τοῦ σημαίνοντος τὸ ὑπάρχω γίνεται τὸ πρῶτον πρόσωπον τῶν πληθυντικῶν ἐμὲν, ἐπειδὴ τὸ πρωτότυπον τῷ ε παραλήγεται, ἔα γάρ· ὅπερ ἐμέν κατὰ πλεονασμὸν σ γίνεται ἐσμέν· δύναται δὲ τὸ ἐσμὲν ἀπὸ τοῦ ἐσμὶ εἶναι·[41]

Here, Choiroboskos not only determined which form is the historically older one but also ventured at a reconstruction **esmi* for the first-person singular. The alternation of *s* and Ø was no longer productive at that time but would have still been clearly recognizable to speakers. Speakers generally attribute morphophonemic alternations

[41] CHOIROBOSKOS 355. Translation: “And from *εἰμι eimi* — with which I begin to expound — arises the first-person plural *ἐμεν emen*, since the original form had an *e* in the penultimate syllable. Because: this *ἐμεν emen* becomes *ἐσμεν esmen* with the superfluous *s*; *ἐσμεν esmen* can be derived from *εσμι esmi*.”

to an underlying phoneme. A speaker of German does not need any historical knowledge to abstract the wordform [hʊnt] 'dog' to the morpheme /hund/. An internal reconstruction can only be assumed when this abstraction is mapped onto a linguistic-historical dimension. This historical dimension is clearly evident in Choiroboskos' work.

There are therefore many approaches that early linguists used that must be associated with the term "Internal Reconstruction" according to today's understanding. For example, Schleicher's (1871:13) explanation for the nominative singular of PIE **mắtēr* 'mother' from **mắtar-s*[42] can be attributed to this type of reconstruction (as done by Joseph 2010:58). Nevertheless, the method only gained authoritative importance with the discovery of the regularity of sound changes. Just as the Neogrammarian hypothesis was previously able to constrain the multitude of possibilities offered by the Comparative Method before, the regularity of sound changes was able to make IR methodologically manageable and to enable reconstruction beyond mere automatic alternation.

5.2.2 Internal Analysis

The term "internal analysis" is used in this thesis to refer to a number of inferring methods for which in a comparative sound correspondence *a:b*, a secondary development in one of the languages is inferred for reasons that can be determined in that language (*langue*). The literature speaks of "internal analysis" (Bonfante 1945:144), "internal reasons" (cf. Porzig 1954:56), or "internal comparison" (Jakobson 1975:506), without defining these terms concretely. In German, the verb *enthaupten* 'to decapitate' is not directly derivable from the word *Haupt* with the meaning 'main' but from the secondary meaning 'head'. It can be inferred from this that the English-German cognate pair *head:Haupt* had the original meaning 'head' and not 'main'. The comparative aspect shines through in this example since purely internally, a development 'main' > 'vital' > 'separate vital' = 'decapitate' also seems possible in theory. These "internal analyses" also include the work of Ascoli (1872), which falls under the term IR according to Watkins (1978:65). Ascoli (1872:5) studied morphophonological alternations but presupposed already comparatively reconstructed sounds in his approach.

[42] Nowadays, one would reconstruct **mắter-s* (cf. Tichy 2009:39).

5.2.3 Structural Method

The term “structural method” (*méthode structurelle*) was coined by Hjelmslev (1966:164) for the method used by Saussure (1879) in his theory of *coefficients sonantiques*, which had been further refined by Møller (1917) to form the so-called laryngeal theory. The derivation of the coefficients was done comparatively (see Latta 1978:64–68). Saussure compared the *e*-grade and zero-grade of Proto-Indo-European verbs with lengthened root vowel and was able to determine two types (Tab. 5.B).

Table 5.B. The *e*-grade of the PIE roots **stā* ‘to stand’ and **dō* ‘to give’ contain a lengthened vowel that, in zero-grade, changes to *a* and *i* in Latin and Vedic Sanskrit, respectively. These *a* or *i* seem to cause the lengthening of the *e*-grade vowel.

	***e*-grade**	***zero*-grade**	***e*-grade**	***zero*-grade**
Vedic	tí-*ṣthā*-mi	*sthi*-táḥ	*dā́*-nam	*dat*-táḥ
Latin	*stā*-re	*sta*-tus	*dō*-num	*da*-tus
PIE	*stā-	*stA-	*dō-	*dǪ-

Saussure explained the lengthened root vowel as *e* + *coefficient* and was thus able to integrate it into the regular *ablaut* scheme seen in Tab. 5.C.

Table 5.C. The Proto-Indo-European ablaut scheme according to Saussure (1879:135). The coefficients behave like semivowels and sonorants.

	base	+ semivowel		+ resonant		+ coefficients		
e-grade	e	ei̯	eu̯	en	em	er	eA > Ā	eǪ > Ō
zero-grade	Ø	i	u	n̥	m̥	r̥	A	Ǫ

This approach can be described as “comparative-structural”: the developments in the different daughter languages (see Tab. 5.B) as well as the proto-language’s root structure (see Tab. 5.C) are taken into account. In addition, Saussure also developed a purely internal approach to explain the irregularity of verbal roots in Sanskrit. The past participle of the Sanskrit verb *man* ‘to think’ is regularly *ma-ta-* (with *ma* as zero-grade of *man*; cf. PIE **mn̥* > Skt. *ma*) and the agent noun *man-tá-r*. In contrast, there are verbs such as *jan* ‘to generate’, which form the corresponding forms *jā-ta-* and *jani-tár*. The ancient Indian grammarians (cf. PĀṆINI 1.2.18 and 3.1.45) called the former roots *aniṭ* (literally “without *i*”) and the latter *seṭ* (“with *i*”). Saussure focused on the nasal presents, which regularly insert an *n* into the present forms before the closing consonant: *yuj* 'to join': *yuṅk-té* (3 sg pres. middle) versus *yuk-tá-* (past participle). Nasal presents also existed among *seṭ*-roots, such as *pū* 'to cleanse': *punī-tḗ* (3 sg pres. middle) versus *pū-tá-* (past participle). Since the nasal is only infixed before the closing consonant, the regular present form would be expected to be ****pnū-tá** or perhaps ****pūn**. The form *punī-tá-* can only have arisen from a root **puX* — with *X* as a *coefficient*.

Both the number of *coefficients* or laryngeals and their phonetic value cannot be determined by IR alone. To determine these, Saussure had to rely on the different phonetic continuations of his *coefficients*. Nevertheless, Saussure's approach already demonstrated the basic features of IR: synchronicity, limitation to one language, and diachronic projection (i.e., reconstruction). It is therefore not wrong to consider laryngeal theory as a result of IR and Saussure as its founder (as done in Lehmann 1962:166, Birnbaum 1970:97, and Hyllested 2009:111). Although his method was not strictly based on the principles of IR, his approach works mainly with purely internal evidence.

Hjelmslev (1966:164), calling the method "méthode structurelle," came to another conclusion. To him (1966:166), a formula such as Saussure's *oA* is not motivated by functions of elements existing in the Indo-European languages but by an internal function of the proto-language. In Saussure's approach, therefore, it was not the "internal structure" of Sanskrit that was grasped but that of Proto-Indo-European. Methodologically, the "structural method" and IR do not differ; they are thus equated by Anttila (1968:166). Only in dealing with external data does Saussure's method seem to be less strict; his second analysis may have limited to Sanskrit for more practical reasons. However, if one follows the view of Hjelmslev, the differences between both methods are much greater. The structural method is applied to a proto-language only *after* the reconstruction and is consequently not a reconstruction method, not even for Pre-Proto-Indo-European. The extensive limitation to Sanskrit is only due to its archaistic structure.

5.2.4 Review

It is difficult to fully agree with Hjelmslev's postulation of a "structural method" since a reconstruction was indeed carried out by Saussure, from Sanskrit to Proto-Indo-European. There are some elements peculiar to the structural method that distinguish it from later concepts of IR: (1) a strict separation between external and internal data is absent, as well as (2) a strict "synchronic" derivation. Instead, there is a loose interdependence between the language in question (Sanskrit), the comparatively inferred proto-language (Proto-Indo-European), and the sister languages (Latin). This interaction is related to (3) the claim of proto-language reconstructions. The reconstructions are attributed to the period of Proto-Indo-European, not to Pre-Sanskrit. As a result, Saussure is inevitably forced to take into account the sister languages, either directly or indirectly.

Consequently, Saussure had no reason to consider his approach to be "internal." He probably saw in it a "special" form of the Comparative Method, which — similar to quite a few reconstructions of Proto-Indo-European — is based on Sanskrit. What is special is that the reconstructions are not based on sound correspondences but on structural analyses.

5.3 Neogrammarian Internal Reconstruction (1907)

Hermann is considered to be the first to systematically describe IR as a method with his 1907 essay (e.g., Szulc 1987:18, Hock 2013:2). He and later Indo-Europeanists were aware that reconstructions based on internal data had already been made. Hermann (1907:16) himself saw his work as being based on that of Wackernagel, only with the additional requirement that, first of all, reconstruction should be done for the individual languages *as far as possible*. His approach is occasionally considered Neogrammarian (cf. Marchand 1956:245) and is thus described as "Neogrammarian IR" in this thesis. Despite its late publication, this designation may well be considered accurate because his paper should be seen in the context of the Neogrammarian discourse. The question of the legitimacy of linguistic reconstruction and the different answers of contemporary linguists led Hermann to the conclusion that the Comparative Method had to be readjusted.

On a formal level, the degrees of uncertainty of reconstructions must be labelled. He (1907:20 and 62) therefore proposed a more complex catalog of labels that would go beyond a simple Schleicher's star to mark reconstructed wordforms. In particular, however, he believed that the Comparative Method was prone to error. Its insufficiency can be reduced with the help of the — in his eyes — more reliable internal-comparative method: From each individual language, one must intend to delve into an older language stage (Hermann 1907:15–16) and can subsequently continue the reconstruction comparatively. Preliminarily, one obtains many "proto-dialects" (*Urdialekte*) instead of one proto-language and each sound that cannot be inferred from the "dialect" must be explicitly characterized as more uncertain. Hermann (1907:16) proposed to name these proto-dialects with the prefix "pre-proto-" (*vorur-*) such as "Pre-Proto-Greek" (*vorurgriechisch*).

5.3.1 Method According to Hermann

Hermann (1907:16) demanded that internal reconstructions must precede each type of reconstruction. Only these reconstructed proto-dialects could serve as a starting point for comparison. In the comparative step, everything that was confirmed by other languages could be reconstructed. According to his examples, he (1907:17) argued for sound distributions and morphological alternations as the basis for IR. He made two assumptions for this:

- A complementary distribution must exist in one environment (e.g., missing final *-m* and frequent final *-n* in ancient Greek).
- These complementary sounds have to alternate in some word pairs (e.g., *χθών chthṓn* 'soil, earth' : *χθαμαλός chthamalós* 'low, near the ground', *ἕν hén* 'one [ntr.]' : *μία mia* 'one [fem.]', *ἔνδον éndon* 'in, within' : *δόμος dómos* 'house').

The hypothesis of etymological allomorphy applies here, as shown in the examples, not only to paradigms but also to semantically similar concepts. Hermann did not draw a firm boundary for what is permissible in non-paradigmatic internal cognates. When he (1907:54) associated Latin *fulvus* 'brown-yellow' and *helvus* 'honey-yellow' and named this as one of two pieces of evidence for the alternation of *f* and *h*, the reader was left with the impression that the sound correspondence was set before the cognate pair.

From these premises Hermann (1907:17) concluded three hypotheses: (1) *-ν* comes from *-*μ*, (2) *μ- (-μ-)* comes from **ν*, or (3) both are created side by side from *x* or *y*. In the next step, he reduced these hypotheses — as in other reconstruction methods — through inferring methods. Hypothesis 2 was rejected since there was no way to explain the words with the preserved **ν* and he succeeded in rejecting the third hypothesis by adding external data.

Hermann suggested in several Indo-European examples that the method of IR was also suitable for reconstructing synchronically less transparent sound changes, such as the Proto-Indo-European syllabic sonorants (cf. Hermann 1907:21–27). However, these were often only traceable in few individual language branches. He devotes a large part of his paper — 29 pages in fact — to the question of the Proto-Indo-European guttural series. In contrast to the majority of his contemporaries, he wanted to assume only two instead of three guttural series. Hermann (1907:58) interpreted the lack of internal evidence for three series as an argument against the majority's opinion. The question about the number of Indo-European guttural series has not been clarified even now, but with Luwian, at least one language is attested today which may have indeed preserved all series for the voiceless gutturals although borrowing may have also been possible in principle (cf. Kümmel 2007:212).

5.3.2 Characteristics of the Method

Hermann's method was based on the morphophonemic and morphological-distributional methods of IR. He described for the first time the formalities of the morphophonemic approach. These included the inferring phase and the hypothesis of etymological allomorphy, which was conceived as a precondition of the relatedness of the words in question (Hermann 1907:19). However, he did not go into detail on problematic issues, such as the determination of pre-sounds and conditions or the reconstructability of different sound change types. And he did not have to since he did not consider IR in isolation from the Comparative Method and thus could resort to external data in case of ambiguities.

In addition to the morphological and distributional approaches, Hermann (1907:31) also reflected upon the possibility of the method of frequency (see Sect. 4.2.2.2) when he considered the lack of Indo-European *b and the dominance of *b^h. However, he (1907:32) was accurate in arguing that the multitude of possibilities did not allow a decision to be made.

However, Hermann incorporated not only internal reconstruction in the narrow sense, but also all reconstruction possibilities that do not require a genetic-external language comparison. This view only became apparent in his last example of the reconstruction of Proto-Indo-European gutturals. In this reconstruction, he implicitly used several other methods:

- SYNCHRONIC-COMPARATIVE METHOD:
  The different pronunciations in modern dialects were taken into account (1907:35).

- External Method:
  The sounds in loanwords were compared with the words in the borrowing language (1907:36).
- Philological Method:
  The historical sound development in the different language stages was included (cf. 1907:37).
- Consideration of attested pronunciations:
  The pronunciation of Sanskrit words was handed down by ancient grammarian (1907:33).

Hermann did not use the term "internal" for this type of reconstruction. He spoke of "Einzelrekonstruktion" (*single reconstruction*, 1907:62) and "Einzelsprachen-Rekonstruktion" (*single-language reconstruction*, 1907:63). Through the term "Einzelsprache" (*single language*), however, Hermann understood "language branch-internal" rather than "language-internal" since the consideration of sister languages was possible in his analysis (e.g., in 1907:54). Only the genetic-comparative method is not used in it. Hermann's approach was lacking in structuralist concepts such as synchronicity or language system, so "Neogrammarian IR" was still clearly different from the later forms.

5.3.3 Review

To Hermann, IR was the preferred method for reconstruction, an attitude that emancipated this method and made scholars aware of it as its own reconstruction method for the first time. However, his proposed method failed to gain acceptance in this particular form. Meillet (1925:12) and Bloomfield (1933:318) ignored his method in their works, writing that the Comparative Method is the only reconstruction method available. One reason for this may have been that he did not succeed in producing new insights and results with his approach. His 64-page paper identifies several examples of IR of Indo-European phonetic questions on about 46 pages, showing that the method reaches the same result as the Comparative Method. Only in the question of the Proto-Indo-European gutturals does he deviate from the opinion of the majority, but here, he is not able to resolve the question internally. Thus, it is not surprising that Reichelt (1922:40) concluded that Hermann, instead of leading and encouraging, had acted as a deterrent, especially since he had not gone much beyond criticism and had relegated the solution of the question to the realm of glottogony. Hermann's basic demand for comparison of exclusively internally reconstructed languages was also demanded in a similar way later, mostly independently of him, such as by Prosdocimi (1977:95) or Chafe (1959:494–495).

5.4 Areal Linguistic Internal Reconstruction (1930s–1940s)

5.4.1 Inner Reasons and Internal Analysis

Without directly building on Hermann's method, the topic of IR gained importance within Italian areal linguistics. The starting point was the *Introduzione alla neolinguistica* by Bàrtoli (1925), in which he addressed the question of which word is to be regarded as inherited when two daughter languages have different continuators. Bàrtoli set forth four principles for this:

- RULE OF THE SUPERIMPOSED LAYER (*norma della fase soprafatta*):
  Of the two stages, the one that is considered older is the one that has disappeared completely or is almost out of use.
- RULE OF MARGINAL AREAS (*norma delle aree laterali*):
  Of the two stages, the one that is considered older is the one that occurs in the peripheral areas.
- RULE OF THE LARGER AREA (*norma dell'area maggiore*):
  Of the two stages, the one that is considered older is the one that occurs in the larger area
- RULE OF THE MORE REMOTE AREA (*norma dell'area meno esposta alle communicazione*):
  Of the two stages, the one that is considered older is the one that occurs in an area far from traffic.

Criticism of Bàrtoli's principles has been raised several times (cf. Pisani 1940:165–167, Porzig 1954:56–58). IR, which Bàrtoli himself does not discuss in detail, plays a role in this theory as inferring method: one of two words is to be considered younger if there are "inner reasons" (cf. Porzig 1954:56) for it.

Pagliaro (1930:174) spoke of phonetic changes caused by INDIZII INTRINSECI. As an example of such indications, he (1930:174) cited the digamma loss in Attic Greek, which, due to the change of *κόρη kórē* 'girl' from **κόρϝη kórFē* (cf. Ionic *κούρη kúrē*), must have occurred after the change of -rē to -ra but before the change *iē*, *eē*, *yē* to *iā, eā, yā* (as in *νέα néā* < **νέϝη néFē*). Similarly, the forms *μέσος mésos* (< **medʰi̯os*), *τείχεσιν téichesin* (< *τείχεσσιν téichessin*), and *βάσις básis* (< *βάτις bátis*) can be classified as more recent processes after the loss of the intervocalic *-s-*.

The examples of *indizii intrinseci* mentioned by Pagliaro have little to do with IR. He may have implied that there must have been a loss of the intervocalic *s* in Greek, but he argued that this did not occur internally but externally via other dialects (Ionic), proto-lingual forms (**medʰi̯os*), or reconstructed forms (**νέFη*).

Bàrtoli (1935:417) himself later introduced the term L'ANALISI 'INTERNA,' which he did not understand as a reconstruction method but rather an inferring method. From an external linguistic comparison of the Latin senex and the Greek *ἕνος hénos* 'old', he inferred an older **s-* since the sound change *s > h* is the typical direction of development (cf. Sect. 4.3.5.1). The term "internal" therefore refers less to "language-internal" than to "sound-internal" since peculiarities of the sound (articulate or evolutionary) are incorporated into the analysis.

5.4.2 Methods According to Pisani and Bonfante

It was only with Pisani (1938:32) that the term *internal* ("ricostruzione interna") was established as a consolidated *terminus technicus* for IR. In contrast to his Italian colleagues, Pisani seemed to build on Neogrammarian IR by including sister languages and historical documents to the method's definition. In this way, he (1938:33) allowed for internal conclusions about proto-languages, such as the notion that unproductive inflectional types such as Latin *arx* 'citadel, fortress' can nevertheless be seen as inherited from Proto-Indo-European even though the word is not attested in other Indo-European languages.

The most influential effect of Italian areal linguistics was probably brought about by Bonfante (1945). In his paper on reconstruction methods, he tried to both capture the reconstruction methods of the past years and bring forth new approaches. To Bonfante, Bàrtoli's areal linguistics occupied a position in the history of linguistics that should not be underestimated; in his view, Bàrtoli's areal linguistics can even be understood as the actual foundation of historical-comparative linguistics. He criticized the traditional Comparative Method without any areal linguistic reference. When Indo-European studies came to the same conclusion, it had actually implicitly applied the methods of areal linguistics, but "frequently they arrived at wrong conclusions […] because of disregard of the norms of areal linguistics and of their respective hierarchy (or order)" (Bonfante 1945:137–138).[43]

Bonfante (1945:133) distinguished between two different "strictly synchronic" methods: IR, which he (1945:132–133) traced back to Hermann, Sàntoli, and Bàrtoli, and the "method of the anomalous form" according to Meillet. According to Bonfante (ibid.), the latter is a method of inferring older linguistic material from paradigmatic irregularities. Unfortunately, he did not delineate the two term definitionally, so the reader could only make assumptions. Meillet himself (1937:37) saw his "method" as merely a means of proving genetic language kinship and not a method of reconstruction. The examples Bonfante gives of "internal" reconstruction are automatic sound alternations compared to these "anomalous" examples, such as *-m#* > *-n#* in *χθαμαλός chthamalós* 'low, near the ground' : *χθών chthṓn* 'soil, earth'. Both methods fall under the notion of IR according to most of today's definitions. In my opinion, a strict distinction between these two methods is neither possible nor meaningful. A distinction between "weak" and "strong" irregularity may be useful since the latter tends to be much more difficult to use as a starting point. However, this distinction has not become widely accepted. In Milewski's (1973:102) classification, which is visibly based on Bonfante's, he understands the corresponding "method of exceptional forms" to be a different approach from Bonfante's (see Sect. 4.3.1.1).

5.4.3 Review

Although areal linguistics had little influence on the methodological approach of IR, even Hoenigswald (1944:86) emphasized the similarities of areal linguistics and IR because both "use typical results of typical historical developments to reconstruct the developments where they are not known from other sources." However, Bonfante in particular had a role in the subsequent history of IR that should not be underestimated. His influence on other linguists was most clearly reflected by Fowkes:

> *When, in 1945, Giuliano Bonfante's article on Reconstruction and Linguistic Method appeared, the present writer belonged to those who believed that there can be no comparative or historical study without at least two chronological stages, or*

[43] He (1945:138–139) went on to cite misguided reconstructions by Indo-Europeanists. On the whole, his version of Proto-Indo-European looked much more like Latin.

two dialectal stages, of the language involved. Comparison seemed to imply the necessity of the existence of two or more things to be compared. Hence he reacted with initial skepticism to the assertion that of some ten ways of studying the history of a language only there require comparison between two languages or dialects.[44]

Although Bonfante merely recited the internal reconstruction methods of other linguists, Greenberg (1966:508) named him first on the list of linguists working on IR. Naert (1957:1) even referred to the various internal methods as "méthodes bonfantiennes" and saw results of those methods in the reconstruction of preliterate Basque by Martinet (1955:370–372).

Nevertheless, Bàrtoli's conclusions did not fall on fertile ground among the supporters of IR. For Kurylowicz (1964:35), areal linguistic conclusions were only "subsidiary arguments," legitimate only as long as they did not contradict internal evidence — an attribute that completely contradicts Bonfante's claim for Indo-European reconstructions. Bonfante's most relevant innovation is the emphasis on the inclusion of universals and non-linguistic information in linguistic reconstruction methods.

5.5 Structuralist Internal Reconstruction (1940s–1950s)

During American structuralism, IR was decisively influenced by the papers of Hoenigswald. In comparison to his predecessors, he moved to a focused reconstruction of sound change instead of individual concrete wordforms. The phoneme with its allophones and its distribution gained crucial importance in his works. In his and the following works, the concept of morphophonemes according to Sapir (1949) and Bloomfield (1933:210–226, esp. 218–219), which describes the alternance of phonemes (not allophones), played a central role. Associated with morphophonemes are the so-called *morphophonemic rules*, which cause the alternation of the phonemes, such as *d* and *t* in languages with final devoicing. Bloomfield described the *morphophonemic analysis* as follows:

[44] Fowkes (1950:142)

> *The process of description leads us to set up each morphological element in a theoretical basic form, and then state the deviation from this basic from which appears when the element is combined with other elements. If one starts with the basic forms and applies our statements [i.e., morphophonemic rules, A.B., ...] in the order in which we give them, one will arrive finally at the forms of words as they are actually spoken. [original emphasis, A.B.]*[45]

Although Bloomfield does not attribute any historical value to his morphophonemic rules, the methodological similarity of IR and morphophonemic analysis is compelling. The "basic form" corresponds to the historical pre-form and the morphophonemic rules to the reconstructed sound changes. Nevertheless, Bloomfield (1939:106) emphasized that the "basic forms are not ancient forms […] and our statements […] are not historical but descriptive, and appear in a purely descriptive order." This was how Bloomfield and Sapir (1949:48–49) left things, arguing that their method has no linguistic-historical claim, without elaborating on differences in procedure. In contrast to this view, Hockett (1948:123) recognized IR in Bloomfield's work on Algonquian. The symmetry of IR and morphophonemic analysis was well known to the followers of structuralist IR (cf. Borgström 1954:275). Lounsbury (1953:11) even unhesitatingly equated the terms "morphophonemic approach" and "method of internal reconstruction" and only saw a difference in their historical or descriptive purpose. Diachrony explains synchronicity, and the morphophonemic rules are merely "productive" sound changes in the diachronic view.

5.5.1 Characteristics of the Method

The aim of the works at this phase of IR was to describe the process formally. For this purpose, a categorization of sound changes with regard to its internal reconstructability was carried out (Hoenigswald 1944, 1965, Marchand 1956) and a focus was placed on sound changes, partially including their relative chronology (e.g., Chafe 1959:481). The method of IR was treated in descriptive detail for the first time and could "scarcely be bettered" by later linguists (Ringe 2003:245, see Sect. 4.1 and 4.2.1). The formal similarity to the Comparative Method was emphasized

[45] Bloomfield (1939:105–106)

(cf. Chafe 1959:478–479, 494, Marchand 1956:245), and the division between both methods was considered “overstressed” (Hoenigswald 1974:189). On the other hand, its independence from the Comparative Method was further established during this period. Under the influence of structuralist dichotomies, a narrowing of the term “internal” took place since Hermann: the method must be accomplished within the *langue* without consideration of dialects or diachronic aspects. At the same time, the description of the method was largely limited to the morphological-distributional approach, so standard works such as Hockett’s (1958:463) mentioned only (a) “morphophonemic irregularities” and (b) “distributional aspect of its phonological system” as bases for IR. Universalist approaches, such as Bonfante’s, were no longer pursued during this period.

5.5.2 Review

To Fox (1995:145–146), IR with its descriptive and non-historical view was the child of structuralism and its first application could be “attributed to no less a person than Saussure himself” with his laryngeal theory. Despite the aforementioned difficulty in seeing the inventor of IR as Saussure, IR actually first gained prominence in structuralist circles. This is especially true of the American variant, in which alternations occupy a weighted space. The works of Hoenigswald, Marchand, and Chafe are the most cited IR papers and still shape IR today. The method of IR subsequently found its way into various introductory works, often as separate chapters (Hockett 1958, Milewski 1973). At the same time, the foundation for the later, skeptical views that many linguists had on IR was laid during this period. Whereas the Neogrammarian IR was intended merely as a prelude to the Comparative Method, the new independence of the IR put its capabilities and potential to a test that it could not withstand.

Synchronic consideration was not necessary for Hermann and Bonfante but became a central component of IR for the structuralists, so the diachronic aspect disappeared completely in later times. Finally, Rosén (in Kurylowicz 1964:33–34) seemed to support such an approach speaking of two kinds of IR. Other than the synchronic type, there would be another type of IR, which tries to reconstruct intermediate forms from two wordforms of different language stages. Today, this type of reconstruction falls under the term “Philological Method.”

5.6 Skeptical Periods

After establishing IR as a reconstruction method in the standard works and thus giving it increased attention, it seems hardly surprising that in the subsequent period, a whole series of papers appeared which were devoted to the problems arising with IR. One question that kept coming up even among the supporters of IR was whether complete exclusion of external data was even possible. Chafe described the state of the debate as early as 1959: "at the moment it is even possible to doubt whether rigorous internal reconstruction is possible at all" (1959:478, cf. Anttila 1973:319). While Chafe and Anttila primarily tried to further develop the methodology of IR, there were more and more papers that identified the problems with IR and stated a preference for the Comparative Method over IR, without rejecting it completely.

5.6.1 Methodical Criticism (1960s–1970s)

In the first phase of widespread criticism, the criticism was increasingly directed against the general approach of IR as presented by the structuralists. Mehendale (1963:42) noted that some reconstructions of Hoenigswald were only possible under the influence of comparative background knowledge. It is precisely the question of the pre-phoneme of a sound change that could often not be answered purely internally. Mehendale (1963:42) also directly attacked Hoenigswald's procedure, writing that Hoenigswald "probably did not feel it necessary to test his statement by applying the procedure of internal reconstruction outlined by him" (cf. also 1963:42 fn. 4). Mehendale (1963:45), however, did not want to abolish IR as a method altogether but saw it as a preparation for the Comparative Method, helping "to eliminate some of the problems of comparative reconstruction." IR thus resumed the position it held before the structuralist phase.

Miranda (1975) again addressed some of Mehendale's points. On the one hand, he (1975:292–294) treated the general question of the validity of the basic hypothesis or the sense of a reconstruction apart without language stages. However, his rather accurate considerations (cf. Sect. 6.2) omitted a comparison of the same questions with Comparative Reconstruction and general reconstruction problems were transferred to IR in isolation. On the other hand, Miranda (1975:294–303) highlighted the problem with determining the direction of sound changes, which was widely underestimated in structuralist IR (cf. 1975:299 fn. 11).

5.6.2 "Generative" Criticism (1970s)

In the structuralist phase, the strong similarity between IR and morphophonemic analysis was recognized and addressed (cf. Zaliznjak 1964:51–52). However, it was not seen as a problem (e.g., Lounsbury 1953:11) since morphophonemic alternations are largely of diachronic origin and the methods differ in their purpose (cf. also Lass and Anderson 1975:117). This non-conflicting attitude changed with the increasing establishment of generative phonology. In contrast to structuralist phonology, generative phonology attaches much greater weight to the process of change of the "underlying forms" (cf. Mayerthaler 1974:1). Generative phonology is also associated with a number of new theoretical approaches, as well as the assumption of a "psychological reality" (Linell 1974). Formally, the "phonological rules," in the same way as the morphophonemic rules, are inferred from productive alternations. For instance, two different "underlying forms" can be postulated for the German "surface form" [bʊnt]: (1) a paradigm /bund/ + the rule [+obstruent] > [- voiced] / _# meaning 'federation' and (2) a paradigm /bunt/ 'colored'. Since the morphophonemic rules had been understood in a purely descriptive way, several alternative rules were, in principle, conceivable on an equal footing — a difference from IR in that it claims that a historical reality can only identify one "rule" as correct. In contrast, phonological rules are considered "mentally constructed" (Chomsky and Halle 1968:14) and thus claim a psychological reality. Consequently, generative phonology considers only one rule to be correct, as IR does so in this respect as well. The generative grammarians had been aware of the parallelism of both analyses (cf. also Halle 1961:89–94):

> *If we now assume that rules are added always singly and always at a given spot in the grammar, then it follows that the synchronic order of the rules will reflect the relative chronology of their appearance in the language. Moreover, under this condition the proposed simplicity criterion can be used as a tool for inferring t history of the language, for it allows us to reconstruct various stages of a language even in the absence of external evidence such as is provided by written records for*

> *by borrowings in or from other languages. It seems to me that such an assumption is made in many studies in historical phonology.*[46]

Nevertheless, the generative phonologists were subsequently accused of merely doing IR or historical-comparative linguistics under a different name (see Derwing 1973:113–115, Maher 1977:11–14, Bailey 1969:85).[47] Supporters of generative grammar, in turn, reproached historical linguistics for doing synchronic analysis. Although most phonological rules have a diachronic background, this fact does not necessarily apply to all rules (for examples, see Sigurd 1966:35). This circumstance resulted in the realization — portentous for IR — that the results of the analysis (IR or morphophonemic analysis) were in fact closer to synchrony than to diachrony, casting "doubt on the credibility of internal reconstruction" (Zeps 1969:150). Sigurd (1966:47), however, emphasized the linguistic-historical value of morphophonemic rules but saw the question of how to distinguish between rules of historical and non-historical origins as a "crucial question" for IR. As a subsidiary solution, he (1966:48) proposed the rule of thumb that the phonological rules "that look like sound-laws might reflect sound-laws."

Among the sharpest critics of this period was Lass (1975). To him, IR is not even part of linguistics but is to be classified as a part of "paleolinguistics" (1975:12) since it reconstructs alternation-free and pure-regular "protolanguages [sic!]" (cf. 1975:19). He illustrated his assessment that the method should be understood as generative rather than diachronic with the nasal sounds in modern French. In French, there are alternances of nasal vowels and nasal consonants, such as in /*nɔ̃*/ 'name' ~ /*nɔme*/ 'to name', /*fɛ̃*/ 'end' ~ /*finir*/ 'to finish', and /*œ̃*/ 'one (m)' and /*ynə*/ 'one (fem.)'. From this alternance, he concluded the diachronic development seen in Tab. 5.D.

[46] Halle (1964:347)
[47] Sometimes, the reader may get the impression that a historical form is targeted as "underlying form." Complex forms, such as in King (1971:211 and 221), are difficult to derive synchronically for a native speaker. In King's example, the derivations are too elaborate for a "mental reality" and the linguist rather used IR.

Table 5.D. Assumed diachronic development for the alternation of *nɔ̃* 'name' : *nɔme* 'to name' and *fɛ̃* 'end' : *finir* 'to finish' according to Lass (1975:13-14). Nasal vowels are derived from nasal consonants.

	**nɔm#*	*nɔme#*	**fin#*	*finir#*
(1) Nasalization	**nɔ̃m#*	-	**fĩn#*	-
(2) Nasal deletion	*nɔ̃#*	-	**fĩ#*	-
(3) Lowering	-	-	*fɛ̃#*	-

Table 5.E. Historically correct development of the corresponding words of Tab. 5.D (Lass 1975:14). These forms are attested or derived from historical documents.

	**nɔm#*	*nɔme#*	**fin#*	*finir#*
(1) Nasalization 1	*nɔ̃m#*	*nɔ̃me#*	-	-
(2) Nasalization 2	-	-	*fĩn#*	*fĩnir#*
(3) Dentalization	*nɔ̃n#*	-	-	-
(4) Denasalization 1	-	-	-	*finir#*
(5) Denasalization 2	-	*nɔme#*	-	-
(6) Nasal deletion	*nɔ̃#*	-	*fĩ#*	-
(7) Lowering	-	-	*fɛ̃#*	-

However, these reconstructions did not fit with the historically correct development shown in Tab. 5.E. IR failed to reconstruct (1) reversed sound changes (e.g., nasalization 2 and denasalization) and (2) merging sounds (e.g., dentalization). This was how Lass concluded that "internally reconstructed" rules are purely synchronic rules and can accordingly be regarded as phonological rules of generative phonology:

> *I conclude that IR is not an independent source of knowledge of language history, since the strong assumption that alone would make it one is untenable. In no case can it be said to yield evidence of an antecedent language state unless there is also independent evidence for that state: in which case IR doesn't 'yield' anything.*[48]

[48] Lass (1975:12)

However, the conclusions that Lass drew in his paper may have been too far-reaching. IR is able to reconstruct conditioned merging fairly well. The dentalization *m* > *n* in his example for a conditioned merger cannot be reconstructed, only because the sound has developed further to Ø, making it impossible to determine whether there was an intermediate step (2a in Fig. 5.1) or not (2b).

In the case of *m* > *n*, case (2a) was accurate, so, in fact, both (1) and (2) in his examples reflect the same problem of non-reconstructability of intermediate steps. This "deficiency" was also characteristic for the Comparative Method and could only be resolved by the Philological Method. Comrie (2003:246), in turn, criticized Lass' requirements for reconstruction. He argued that the final reconstruction **nɔm* had coincided with an actual historical form. The pre-form itself had said nothing about its historical development. Neither the Comparative Method nor IR had a claim to reconstruct a complete phonetic history but only to reconstruct a proto- or pre-language or parts of it.

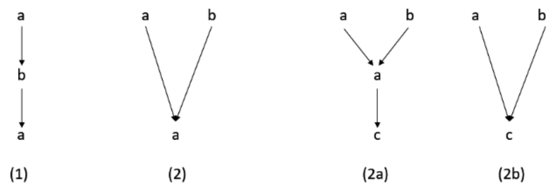

Figure 5.1: Problematic Sound changes for IR according to Lass (1975:14): (1) reversed sound changes and (2) mergers. In fact, IR cannot differ between merging with (2a) and without intermediate stage (2b).

Notwithstanding, Lass (1975:15) concluded that IR "cannot be a historical reconstruction: so the only other thing it can be, apparently, is a synchronic description," ultimately arguing that IR, in fact, IS generative phonology. However, he (1975:17) stated that not every morphophonemic alternation was "psychologically real" either, so generative phonology "under current definitions is synchronically false" as well.[49] Both were "abstract morphophonemic analysis" (1975:19). Whereas generative phonology had "considerable value" for him as a cognitive theory, IR was "of no use at all and might as well be scrapped" (ibid.).

[49] As an example, he (1975:15–17) cited the active and passive forms in Maori, such as *awhi* 'embrace' : *awhitia*, *hopu* 'catch' : *hopukia*, *aru* 'follow' : *arumia*. The diachronically correct passive form is formed with *-ia*, but the active form has a consonant loss in final position. Therefore, speakers tend to regard *-tia* as passive suffix.

In this phase in which many scholars were critical of IR, IR's raison d'être was called into question for the first time, forcing contemporary supporters of IR, including Anttila and Milewski, to reevaluate the method. In the long run, however, this extremely skeptical view on IR did not prevail. Instead, one did not see any problem in the methodological homogeneity of both methods because there is no reason "why similar theory and methodology cannot be applied to diachronic and synchronic analysis alike to bring the current schizophrenia in linguistics to an end" (Bailey 1969:105). Even Lass (1997:232–241) evaluated IR much less skeptically in his later works, such as his 1997 introduction to historical linguistics.

5.6.3 Indo-Europeanist Criticism (1970s–1980s)

While the topic of IR had seemed to be of little relevance to Indo-Europeanists in the period between Hermann and Kurylowicz, it increasingly became the subject of discussion in the 1970s in the context of the establishment of laryngeal theory. Boretzky, an opponent of the laryngeal theory, argued in his 1975 paper that laryngeal theory, which he (1975:47) considered an application of IR, did not actually behave as "normal" IR (1975:57) did and consequently was methodologically fragile. The laryngeal theorist Mayrhofer (1980:364), on the other hand, saw the laryngeals as a product of most extreme comparative reconstruction and not a result of IR. Another position was taken by Gmür (1986:41), who saw the laryngeal theory merely as the application of the method of analogy.

The internal approach of the laryngeal theory has already been explained in Sect. 5.2.3 and can be called internal reconstruction in its procedure, although this approach had been preceded by a comparative reconstruction and the purely internal method would not have led to today's trilaryngealism. Its procedure also corresponds to Boretzky's (1975:50) definition of IR, working with alloforms in a purely synchronic way. His criticism of laryngeal theory is also not justified insofar as he did not address the derivation of the laryngeals and their evidence but limited the laryngeal theory to its consequences for the Proto-Indo-European root system. For example, he (1975:57–58) criticized that one would also have to conclude from the sequence **#HC-* (laryngeal + consonant in initial position) a sequence **#RC* (resonant + consonant in initial position). The actual reason for the reconstruction **#HC* remained unmentioned. Boretzky (1977:30) concluded that Saussure's laryngeal theory is primarily an attempt to regularize the proto-language. He ultimately criticized the results of the laryngeal theory, not its method. In contrast, Mayrhofer's

(1982:177) view felt short of trying to prove that the laryngeal theory is not an application of IR:

> *[D]ie Gedanken in SAUSSURES „Mémoire“ sind keine interne Rekonstruktion (selbst wenn wir die Möglichkeit interner Rekonstruktion in erschlossenen Sprachen nicht prinzipiell ablehnen), sondern ein induktiver Versuch, die indogermanische Wurzelstruktur besser zu verstehen, der zu seiner Zeit wohl noch nicht verifizierbar war.*[50]

Mayrhofer (1982:178) may have at least been right in the fact that the laryngeal theory cannot be derived only from the Avestan alternation of nom. sg. *paṇtā̊* : gen. sg. *paθō* ‘path’ although it was caused by laryngeals. However, IR here did not lead to the assumption of “irregularity,” (i.e., a suppletive paradigm) but to several possibilities of reconstructions.

Mayrhofer (1982:178 and 181) considered laryngeal theory the “most extreme” Comparative Method, which only shares with IR the desire for uniform lexemes. The “external” part results from the fact that, unlike IR, it is based on the sum of the experience of language comparison. Taking this idea to its logical conclusion, even the basic hypothesis of IR would ultimately be an “experience” of historical linguistics made in other languages. Linguistic reconstructions are fundamentally based on the assumption that a language behaves like any other language.

5.7 Universalistic Internal Reconstruction (1960s–1970s)

In the works of Kurylowicz (1964 and 1973), a new approach to IR began. In terms of the history of science, his paper can be assigned to structuralism with a clear tendency toward universalistic research, while displaying the influence of the works

[50] Mayrhofer (1980:365). Translation: “The thoughts in SAUSSURE’s “Mémoire” are not an internal reconstruction (even if we do not reject, in principle, the possibility of internal reconstruction in reconstructed languages) but an inductive attempt to better understand Indo-European root patterns, which was probably not yet verifiable at his time.”

of Hoenigswald and Martinet (cf. Kurylowicz 1964:10–12). With the increased integration of typological and universalistic research, he followed a scientific trend that had been established in historical linguistics in the post-war period (cf. Birnbaum 1977:20–21). While historical linguistics of the pre-war period was strongly fixated on reconstructions, an increased interest in "diachronic change itself" began after World War II (Givón 1999a:109). The integration of typology was mostly carried out in a regularizing manner, in which the result of the reconstruction had to be corrected with the help of typological linguistic laws (see Jakobson 1958:23). As a result, new theses about Indo-European vocalism or glottal theory arose. In contrast to these works, Kurylowicz tried to integrate typology more into the reconstruction method itself. Birnbaum described Kurylowicz' view as follows:

> *Kurylowicz seek to define the particular nature of internal reconstruction as a new linguistic term, endowed with a special meaning only a posteriori — in other words, he contrasts it with linguistic construction in the traditional sense, i.e., reconstruction utilizing data provided by the comparative method. According to Kurylowicz, we must interpret internal reconstruction as designating specifically* LINGUISTIC *reconstruction in contradistinction to other approaches aimed at the recovery of phenomena and processes not entirely of a linguistic nature and operating, therefore in conjunction with extra-linguistic methods as well. Kurylowicz points specifically to the "chronological order of the reconstructed linguistic facts (relative chronology)" as the goal of all linguistic reconstruction*[51]

5.7.1 Characteristics of the Method

Kurylowicz himself saw in IR great potential for gaining of new knowledge about general linguistics (especially generative grammar and typology). To him, IR offered the possibility of applying empirically acquired findings of historical linguistics (i.e., universals) that should not be underestimated because "general linguistic considerations seem to play an ever increasing role in historical reconstruction"

[51] Birnbaum (1970:96)

(ibid.). In doing so, he revisited Bonfante's idea of using UNIVERSAL RESEARCH for linguistic reconstructions, which was not yet widely developed at that time. With the additional application of diachronic universals, he tried to grasp deeper historical stages, instead of conducting a simple synchronic phoneme analysis. The universalistic approach was continued by Greenberg (1978:75–76), who tried to model it and took into account different types of universals.

Methodologically, Kurylowicz's approaches include a departure from phonetic IR to GRAMMATICAL RECONSTRUCTION (see Sect. 4.3). In this view, wordforms are only the starting point of analysis; IR itself is "heavily grammar-oriented, and thus concentrates on the subsections of grammar (and vocabulary) rather than on individual words" (Anttila 1973:330). Kurylowicz seemed to want to counter the descriptive nature of structuralist IR, which hardly went beyond the morphophonemic method. Nevertheless, universalistic IR found few imitators. Among them, the most notable was Anttila, who, in contrast to Kurylowicz, considered "non-internal" knowledge (including extra-linguistic information) to be necessary in IR; otherwise, only an analysis has been conducted. According to Anttila (1973:324), IR that considers only the system of language and leaves out other facts, such as dialect geography, areal linguistics, or glottochronology, imposes demands on itself that are impossible to meet "since one has to incorporate semantics with its 'extralinguistic' connections as well as intralinguistic variation." The integration of dialects brought IR back toward the Comparative Method, which Anttila considered "identical" to IR. With the integration of linguistic knowledge, the method distanced itself from phonemic analysis (Anttila 1973:321).

Kurylowicz's definition of RELATIVE CHRONOLOGY (i.e., the chronological order of the reconstructed linguistic facts) found far more consensus among his contemporaries. Relative chronology was defined as the "prime task" of IR (Anttila 1973:325) or even of all linguistic reconstructions (cf. Birnbaum 1970:96). This development can be seen as a reaction to the problems of wordform reconstruction that were increasingly made clear in IR-skeptical literature. Anttila (1973:344 and 340) therefore stated that IR "cannot reconstruct, but we get relative chronology of types." By the term "types," he meant to indicate alternation and inflectional types. In this context, IR as a reconstruction method was redefined as a pure method for determining relative older and younger elements of a language. With this redefinition, a divergence took place between IR, which was regarded as formally similar to the Comparative Method during the structuralist phase, and the Comparative Method itself. While IR determines the older of two elements, the Comparative Method "aims at determining which of the elements of different, related languages is older" (Milewski 1973:103).

5.7.2 Review

The two most important characteristics of universalistic IR — universals and grammatical reconstruction — found few imitators in later periods, such that even Campbell and Grondona (2007:24) saw Kurylowicz's IR as a method that had little to do with IR today. Many linguists did not seem "too sanguine" about the use of universal sound laws because of a lack of solid examples (Hoenigswald 1974:188–189). In Kurylowicz's grammatical reconstruction, he repeatedly resorted to other languages (see Sect. 4.3.3), thus fulfilling his extralinguistic condition but not making clear the restriction to a single language (cf. Birnbaum 1970:96 fn. 8 and Hock 2013:5). Greater importance was attached to relative chronology, which, in principle, stemmed from structuralist IR. With the relative chronology, the adherents of IR succeeded in countering some criticisms of morphophonemic IR, but it also became the method that was restricted to this issue. As a reconstruction method, it thus became increasingly less important and was generally classified as a method of "secondary significance" (Milewski 1973:110). Even though IR seemed predestined to be the future method of historical-linguistic reconstruction to Kurylowicz, even his supporters largely did not follow this assessment. The Uralist Anttila (1973:349) ultimately concluded that IR "is not powerful enough for lifting its results as the 'new truth' in Uralic studies."

5.8 Modern View

From the 1980s on, the dispute about the homogeneity of IR and phonemic analysis began to subside. Although it was still considered a problem (cf. Fox 1995:210), an acceptance of both methods could be established, ultimately also because of their different purposes:

> *[S]ynchronic phonological theory places a high value on productivity, which may in turn be the result of analogical change, whereas internal reconstruction stresses the importance of irregularities, often so rare that synchronic phonologies would merely assign them an exception feature of*

> *some kind. The least productive and most irregular alternations are often the most revealing for the comparative linguist, but the most productive and least irregular alternations are the ones that best serve the synchronist.*[52]

In addition to the different handling of the "irregularities," IR also showed a "slightly different emphasis" (Anttila 1989:265) on the phonetic environments. The acceptance of the method as a reconstruction method led to the reassertion of procedures and goals from the structuralist phase. Relative chronology as a goal of IR continued to lose relevance ultimately because "the cogency of the interferences is often [regarded as, A.B.] overstated" (Hoenigswald 1992:31).

In more recent papers, the notion of the Comparative Method as the "gold standard" (Kiparsky 2014:65) manifested itself, to which IR joined merely as an auxiliary method (e.g., Givón 1999a:107–109, Hajnal 2016:438). This attitude is already evident in standard works. The first page of Ringe's work (2003:244) immediately began with an emphasis on the "less reliable" nature of the method and a listing of its problems. Unlike during the structuralist phase, the differences between the comparative and internal method are emphasized. Working with morphophonemes, IR is "more abstract" and "more grammatical" than the Comparative Method (Fox 1995:213–214). A more significant difference can be seen in the evaluation of the results. IR offers only plausible results (cf. Hajnal 2016:439) and requires "several assumptions about which types of changes are most likely to have given rise to the synchronic patterns observed" (Ringe 2003:244). Ringe incorporated inferring methods under "several assumptions" and he regarded the Neogrammarian hypothesis as the only IR assumption "that is completely reliable in every case" (ibid.). In doing so, he followed the hypotheses-building theory, as supported by Latta or Bonfante. According to this theory, a reconstruction method is to be understood as the sum of its hypotheses (including basic hypothesis and inferring methods; see Sect. 3.1.1.1).

Papers since the 2000s have also been characterized by a certain degree of skepticism toward the Comparative Method, which is based on uncertainties in the determination of a proto-language's phoneme system from only a few individual languages (e.g., Unger 2000:655). If the language family consists of only a few

[52] Rankin (2003:189)

languages or even only a single language, erroneous reconstructions, as in Hermann's thought experiment, are inevitable. In these philologies, IR and Comparative Method have been established as mutual "cross-check" techniques, and IR has developed as the "major theme of research on Japanese and Korean over the past several decades" (Unger 2000:655).

The change in the understanding of IR can be seen in Campbell and Grondona's (2007:24 fn. 15) statement that the papers by Naert and Kurylowicz "actually bear very little connection with the method of internal reconstruction as recognized today." Instead, the structuralist definition increasingly prevailed and, with it, a focus on the morphophonemic method and pattern recognition, both of which are also mentioned in introductory works (e.g., Ringe 2003 and Anttila 1989:264–266). There has been a tendency to move away from the strict "algorithmic definition" (cf. Campbell 2013:199 and 203). In particular, the phonotactic part of pattern reconstruction gained importance in some philologies, as alternations are far less common in the indigenous languages of the Americas and Africa (cf. Landerman 1997:35). Other approaches have been increasingly viewed in a critical light. While Hoenigswald considered a frequency-based method on the example for Sanskrit *a* possible, Campbell (2013:209) saw "no basis in Sanskrit itself for seeing anything else in the past of the *a*." However, the use of typological or "plausible" rules and universals has been completely discarded but continues to be used as a corrective of internally reconstructed sound changes — and thus as a method for drawing conclusions.

5.9 Use Cases of Internal Reconstruction

Regardless of the theoretical treatment, the establishment of a reconstruction method is expressed only by its practical application to concrete languages. The separation between a theoretical and a practical implementation is not always clear. Fowkes (1950) already applied internal methods to Welsh, but to him, the focus was more on evaluating the method rather than reconstructing its pre-language. A "practical application" can only be spoken of when a method for gaining knowledge is used.

An early area of application concerned the indigenous languages of North America. The early methodological considerations on IR of Lounsbury (1953) and Chafe (1959) already served to establish IR as a reconstruction method for the Iroquoian languages Oneida and Seneca, respectively. Influenced by the definition of structuralist IR, this phase was heavily morphophonemic (e.g., Hoijer 1969 for Navaho and

Cook 1974 for Sarcee). The languages of North America treated mostly belonged to language families with few individual languages, so IR was intended to facilitate the reconstruction of the proto-language. As early as 1969, Haas summed up that “the principles of internal reconstruction have been used with considerable success” with indigenous languages (Haas 1969:109). The indigenous languages of other continents became the subject of IR in subsequent eras since these mostly later became a focus of historical linguistics (e.g., Gilyak in Siberia by Austerlitz 1982, 1984, 1994, Nilo-Saharan Päri by Andersen 1989, and the Arawa language Jarawa by Dixon 2001).

Other than the morphophonemic method, there were also attempts of a phonotactically influenced method (e.g., Freeze 1976 for Hopi, Andersen 1989) as well as of pattern reconstructions (e.g., Adams 1985 for Mutsun, Michailovsky 2012 for Sino-Tibetan Dumi). Attempts to introduce new approaches come from Austerlitz (1990), who tried to reconstruct the proto-phonemic system using typology. Departures from the morphophonemic method are more frequent because those languages are only weakly inflectional and agglutinative languages usually have fewer starting points for this method.

Languages from smaller language families also dominated the literature of IR in later times. The application of IR to proto-languages accounts for only a minor part (e.g., Schmidt 1979, Lehmann 1989 and 2002, Rasmussen 2009 for Proto-Indo-European, Fronzaroli 1970 for Proto-Semitic), as well as for isolated languages (e.g., Haas 1980 for Hokan group, Austerlitz 1983 for Gilyak, Martínez-Areta 2013 for Basque). Working with isolated languages is often associated with trying to connect them with a language family (e.g., Blevins 2018 for Basque-Indo-European). Generally speaking, there has been a tendency for IR to increase as an area of application in recent years (e.g., Norton 2013, De la Fuente 2017, Jacques 2017, Blažek and Bičanová 2014).

6. Criticism of and Problems With Internal Reconstruction

The changing methodology and definition of IR throughout the course of its history is to be understood as an attempt to do justice to the claim of a non-comparative reconstruction on the one hand and to address ongoing criticism on the other. Despite its long history, it has never gained the importance that the Comparative Method has taken in historical-comparative linguistics since Neogrammarian times. Consequently, it did not succeed in completely eliminating all the remaining problems in its procedure. The advantage of a non-comparative method has been well recognized. The majority of Indo-Europeanists thus accept a Pre-Indo-European connection between accent and *ablaut* although this cannot be confirmed externally. In this chapter, the most important criticisms of IR will be recapitulated and will themselves be critiqued.

6.1 Can Diachrony Be Derived from Synchrony? Criticism of the Diachronic Projection

6.1.1 Synchronic Alternation or Diachronic Sound Change?

Especially at the peak of generative grammar, there was growing criticism that IR — specifically, the morphophonemic method — was more a synchronic description of an alternation or a phonological analysis than a correct historical reconstruction (cf. Lounsbury 1953:14–15, Miranda 1975:304, Ringe 2003:246). The causal dependence of synchronic alternation and diachronic sound change is ultimately the decisive criterion for the approach of IR that is based on alternances. While a variety of morphophonemic rules correspond to historical sound changes, they can also be of **analogical origin**. Lounsbury (1953:15) cited as examples the following morphophonemic rules from the Iroquoian language Oneida (*C* stands for any consonant):

1. h*C*h > *C*h
2. a + i > ʌ
3. ʔ + w > h

Of these three rules, the first is automatic and can be traced back to a historical sound change. The second and third rules apply only to pre-pronominal and pronominal prefixes. The second rule is historically much older than the first one and is probably based on a sound law as well. In contrast to these, the third rule is of analogical origin (ibid.).

A deviation from synchronic and diachronic rules can also be caused by REORGANIZATION OF A SOUND CHANGE (cf. Chen 1976:216). The Gothic sound change *t* > *Þ* \ [-fricative]_ appears synchronically to the speaker as *Þ* > *t* \ [+fricative]_ since speakers tend to recognize positive conditions in phonological rules (for more examples of so-called *rule simplification*, see King 1971:224). In these cases, IR also tends to favor the latter interpretation and thus does not correctly identify the direction of sound change (cf. Sect. 4.2.1.4). Far more problematic for IR are restrictions and extensions of synchronic rules[53] and synchronic rule insertions (cf. Chen 1976:220). For instance, the Slavic sound change *s* > *x / i, u, r, k_* appears synchronically in Old Church Slavonic as *s* > *x / k,g,*V+_(+)V (Zeps 1969:145). The rule was narrowed by the secondary changes *ds* > *s* and **ḱ* > *s*, whereby the sound *s* again occurred in positions after *i* and *u*: **boidsos* > OCS *běsъ* 'devil'. The secondary sound changes cannot be recovered synchronically, so the rule had to be restricted in its condition. The extension of the rule to the environment *V_V* arose by analogy, when the regular alternation of *s-x* was extended from aorist forms of *i*-stems to other vocalic stems as well, and forms such as *glagolaste* 'you (pl.) spoke' : *glagolaxъ* 'I spoke' were created (explanation according to Zeps 1969:145). Extensions and restrictions are generally only internally reconstructable if the forms have left relicts.

Table 6.A. Historical development of Proto-Slavic *wilke 'wulf' and *kainā 'price'.

	**wilke* 'wulf'	**kainā* 'price'
1st palatalization	*vlьče*	**koina*
monophthongization of **oi*	*vlьče*	**kěna*
2nd palatalization	*vlьče*	*cěna*

[53] This corresponds to type B of the Kiparsky's typology (1968).

Furthermore, Zeps (1969:146) postulated an additional discrepancy between relative chronology and the "descriptive order" (another example is cited by Rosen in Kurylowicz 1964:34). Historically, there were two palatalization processes of velars before front vowels in Slavic (see Tab. 6.A). Zeps (1969:149–150) criticized the fact that there are a variety of ways to arrange the rules synchronically. If one is willing to also consider the exceptions (mostly created by analogy), the necessary number of "rules" increases, as well as the possibilities of ordering. Instead of many rules, however, a synchronic speaker tends to see one rule that is *k, g, x* become *c, dz, s* / _+{i,ě}+ and *č, ž, š* / (+)FV.

The rule *k, g, x* > *c, dz, s* reflects the second palatalization but seems synchronically to be the older one; otherwise, *_i/ě* as front vowels would have been affected by this rule. The remarkable thing, however, is that Chomsky and Halle (1968:423–426 and 428–430), in their derivation of the phonological rules of Slavic palatals, came to a different conclusion, more in line with the established relative chronology of the two regressive velar palatalizations. Differences between the generative phonological rules and IR arise, especially in the "ordering" of rules and sound changes. According to Birnbaum (1970:113), the rule ordering for generative rules should be chosen according to the criteria of economy and "maximal simplicity."

From the examples provided, it is clear that the synchronic rules are not identical with the diachronic sound changes, but they are rather the results of them and, as such, correspond to the approximation postulate of reconstruction. The criticism is due from Hoenigswald's limitation of IR to a morphophonemic method. The Comparative Method holds a comparable position in the work of Bailey (1969), who likewise equates the Comparative Method with synchronic "pandialectal phonology" (1969:86).[54] According to this theory, every speaker possesses a "polydialectal competence" (1969:89) with dialectal sound correspondences, which he acquires by mentally applying a kind of Comparative Method. As in the case of IR, the methods are almost identical; the results only "differ because [of] the aims" (1969:87).[55] Where synchronic analysis differs from the diachronic analysis, Bailey (1969:97) argued that it was merely to generalize, or "the analysis may be wrong in that the unnatural rule should be reduced to two or more natural ones." Nevertheless, it must be noted as an objection that the Comparative Method does not consist only of the derivation of sound correspondences but that the application of inferring methods

54 The structural similarity between dialectology and Comparative Method was also observed by Dyen (1963:632–633).

55 However, some slight deviations are accepted, because "synchronic formulas normally show some loss, as compared with the related realities" (Bailey 1969:104).

constitutes an important part of the method. A similar procedure for determining underlying forms may be available synchronically; for example, English dialect speakers who pronounce *taut, tort,* and *taught* homophonously may certainly recognize other underlying forms when they come into contact with "*r*-ful" dialects. However, linguists have far more possibilities available to them from their empirical or theoretical knowledge and may thus arrive at reconstructions other than the underlying form. A similar situation exists between morphophonemic IR and phonological rules. Differences have arisen here mainly due to inferring methods. Thus, in the case of Gothic phonetic alternation, considerations of parallels in the behavior of the other plosives could lead to historically correct results although this is not readily possible in the specific Gothic case. The abstraction to sound laws (see Sect. 4.2.1.6.2) is not necessarily a component of a morphophonemic analysis either. A morphophonemic analysis projects German [t] and word-final /d/ to a synchronic morphophoneme {T}, while a morphophonemic IR does not seem to provide any new information. Moreover, an abstraction to *C[+obstruent, +voiced] > C[+obstruent, -voiced] / _#* or to a kind of abstract morphophoneme {consonant} no longer occurs here.

Such criticism may cast doubt on the reliability of the morphophonemic approach of IR but does not attack its legitimacy in principle. Among the supporters of IR, the view "that historical and descriptive linguistics support each other" has prevailed (Anttila 1968:171). According to Anttila (1973:319), a reconstruction as such, however, requires additional "background" knowledge, such as universals or oppositions, to free IR from its synchronic dilemma. He (1973:323) himself saw here also the knowledge of historical and dialectal forms as legitimate "background" and was skeptical of a purely internal reconstruction.

6.1.2 Etymology or Folk Etymology?

Closely related to the question of the difference between phonemic analysis and IR is the question of the demarcation between reconstruction and synchronic knowledge of diachrony (i.e., is the linguist actually reconstructing a historical pre-form, or is he merely re-producing the picture of diachrony that appears to the speaker synchronically as such?). Linguistic deviation in idioms and in poetically marked expressions are perceived by speakers as "archaic." Even morphophonemic alternations may belong to this "archaic image" of the speaker. Proving or disproving such perceptions may be the real task of a historical linguistic examination.

However, I question here the principal sense of a "reconstruction" on the basis of automatic alternations. If a sound alternation is productive, it is synchronic and does not need to be "reconstructed." That the German wordform [lant] belongs to a morpheme /land/ is known to a native speaker and is expressed accordingly in orthography. If language reconstruction is understood as an attempt to restore older language stages or language history, this form of IR can only fall under this term if one understands language not as *langue* but merely as *parole*, completely setting aside the language system. In this thesis, therefore, the view that only sound changes are admissible as reconstructions if they are no longer synchronically productive is taken. After all, the general principle for diachronic linguistics also applies here: the necessity of finding the conditioning environments that cannot be derived immediately from the data essentially distinguishes between the role of the comparative linguist from that of the descriptive phonologist (Katičić 1966:212).

6.2 Does Internal Reconstruction Consider the Right Assumptions? Criticism of the Basic Hypothesis

6.2.1 Suppletion as an Exception to the Basic Hypothesis

IR is based on the hypothesis that two allomorphs of the same language can be traced back to a common root morpheme. The fact that not all allomorphs go back to the same etymon is well known to the methodologists of IR and is more or less addressed in treatises (e.g., Chafe 1959:480 and Marchand 1956:246–247). Various methods have been proposed for this purpose (e.g., to identify suppletion on basis of phonetic similarity; see Sect. 4.2.1.2). Oftentimes, however, suppletion arises only from the phonetic and semantic similarity of different morphemes. Old High German had two different words for 'to go': *gân* (without any preterit form) and *gangan* (with preterit forms; e.g., *giang* 3rd sg.). Both paradigms merged in later times: as present forms, the wordforms of *gân* prevailed, and as preterit forms, the wordforms of *gangan* prevailed. This development certainly took place under the influence of the sporadic alternance of *h* and *ng* (e.g., OHG *ha**h**an* 'to hang (inf.)' : *hia**ng**un* 'they hung'). The coincidence of the two paradigms happened, so to speak, under "folk-etymological" influence.

Such kinds of folk-etymological suppletion, which also includes lexical suppletion, are difficult to identify internally as well as comparatively and can often only be determined philologically. A comparable problem exists with non-cognate words with similar or identical meaning. Nevertheless, suppletion does not play the problematic role often ascribed to it by critics. In particular, paradigmatic suppletion only occurs in a small proportion of the total vocabulary; the sound correspondences erroneously obtained from it therefore carry little weight. For that reason, it is advisable to consider the relative frequency of sound correspondences when using automated methods.

6.2.2 Is the Basic Hypothesis Correct?

Regardless of suppletion, the question of whether alternations must in principle go back to a regular origin remains. Some scholars, such as Ringe (2003:252), indicated that an alternation could have already existed in the proto-language (e.g., Proto-Indo-European *ablaut* or *r/n*-heteroclitic declination). To circumvent this problem, one must contend with a deeper time level (i.e., one only has to go back far enough in time to achieve the causal sound change; cf. Miranda 1975:292). This line of argument is quite valid since IR, by its definition, does not attempt to reconstruct proto-languages but only elements of the pre-language. Nevertheless, the possibility of a rule that has "always existed" has been raised repeatedly in the literature. King (1971:205) referred to sound changes of universal character, such as the nasal assimilation point rule. Miranda (1975:293) saw as actual cases of invalid basic hypothesis so-called "INVIOLABLE ARTICULATORY" CONSTRAINTS, which refer to articulatory forced sound changes (e.g., the pronunciation of vowels before retroflex consonants as retroflex vowels). However, provided that the pronunciation here is indeed physically forced, these cases do not play a major role in reconstruction in general and specifically in IR. In the simplest case, they can be treated as regular sound changes (e.g., *a* > *a̢* / _C̢), which occur as soon as corresponding sound sequences arise anew in a language.

Lass (1975) had a far more critical view of this question, arguing that the validity of IR depends on the correctness of the basic hypothesis. If it is false, then "IR is in principle impossible, except under very narrowly constrained conditions," and these are "so restrictive as to make IR virtually supererogatory when it is possible" (Lass 1975:9). He himself does not give an example of non-historically conditioned sound alternations, apart from the remark that there is no evidence that the Proto-Indo-European *ablaut* is actually due to a diachronic change (Lass 1975:10). Such examples are rare indeed. Although there are, for example, morphologically or analogically secondary *umlauts* in German, these do not change the fact that the synchronic "umlaut" is due to a historical sound change. A non-historical alternation would only apply if two etymologically unrelated, but phonetically almost identical forms are united into a suppletive paradigm and the difference is understood as an alternation and SPREADS ANALOGICALLY TO OTHER PARADIGMS. Such a case is assumed for Latin *premere* ‘to press’ : *pressisse* ‘have pressed’, which is analogous to an older suppletive paradigm *tremo* ‘I tremble’ : **tress-* (cf. Meiser 2003:116). However, the admissibility of a method does not depend on the invariable correctness of the basic hypothesis. The Comparative Method is admissible even though it is not always certain in individual cases whether a cognate pair represents an inherited wordform from the proto-language or an early loanword. The validity of the results of a reconstruction depends directly on the probability of the basic hypothesis occurring. This probability seems to be high for the basic hypothesis of IR.

6.2.3 The Basic Hypothesis Leads to a Regular Proto-Language

A problem hotly debated until the 1970s was the reconstruction of a pre-language as an alternation-free and consistently regular proto-language (cf. Miranda 1975:293, Lounsbury 1953:14, Lass 1975:10). In principle, IR can only map a partially irregular inflectional language onto a consistently regular inflectional language, as this is the only direction that allows the basic hypothesis (Chafe 1959:495). The consistent elimination of alternations and irregularities “would easily end up with perfect agglutination” (Anttila 1973:346), and the reconstructions could consequently be called “over-simple results” (Lass 1975:15); IR, just as morphophonemic analysis, merely attempts to reduce a large number of variants to one entity.

However, these critics do not take into account that IR reconstructs only pre-languages and no — as Lass (1975:10) puts it — “protolanguages with NO ALTERNATIONS AT ALL [original emphasis, A.B.].” IR reconstructs wordforms that could be assigned to different language stages (see Fig. 6.1). Language history is marked by the emergence and elimination of alternations. IR merely reconstructs the time of the alternation’s emergence, thus inevitably resulting in a purely regular pre-language. This problem was already known to early supporters of IR and led to the relative reconstruction stage “pre-language” with “uncertain and uneven chronological depth” (Anttila 1968:165). Occasionally, a rejection of wordform reconstructions is caused by this problem, as well, because IR “can give the age of the *general types* only, not *individual items* in them [original emphasis, A.B.]” (Anttila 1973:340).

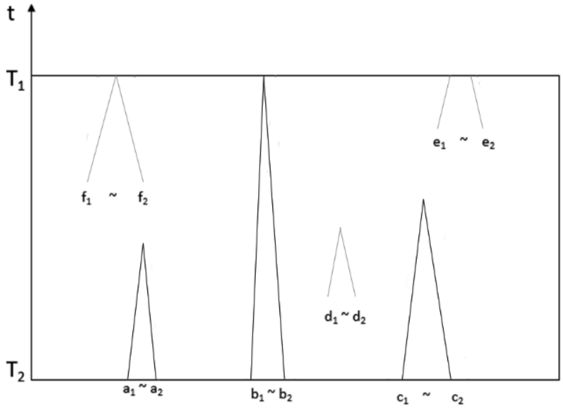

Figure 6.1: Emergence and elimination of alternations. IR only reconstructs emergences (black lines) and cannot state anything about neither time level nor lost alternations. The illustration is loosely based on Miranda (1975:294).

Fox brought a new dimension to the debate. He (1995:211) indicated that the Comparative Method in standard works reconstructs a proto-language that is in many ways richer "than any of the daughter languages themselves." He attributed this, among other things, to the Comparative Method's ability to reconstruct mergers which are mostly identified "when the number of correspondence sets is greater than the number of phonemes in individual languages, and no evidence can be found for the complementarity of the sets" (1995:213). The consequence is that the Comparative Method has a "tendency to over-reconstruct mergers" (ibid.). This aspect may also explain the different views that historical linguists have about the phoneme number of proto-language (e.g., the number of laryngeals or gutturals in Proto-Indo-European) but especially the different reconstructions in the field of grammar. If one of two daughter languages possesses an indicative and optative, the other one an indicative and subjunctive, there is often no possibility to prove a category as a linguistic-historical innovation. The simplest solution would then be a reconstruction of three categories for the proto-language and thus a more complex language. In contrast to IR, however, the linguist has the possibility of marking uncertain reconstructions as such. He is able to mention both possibilities (i.e., two or three categories) for the proto-language. IR does not have such options.

6.3 Are the Results Useful for Historical Linguistics? Criticism of the Reconstructions

6.3.1 The Reconstructions Are Indeterminate in Time

The fact that IR reconstructs only "pre-forms" is linked to a TEMPORAL FUZZINESS or variance of the reconstruction: The Indo-European *ablaut*, which is probably of Pre-Proto-Indo-European origin, is still reflected in some modern languages (see English *sing : sang : sung*), while more recent sound changes can often no longer be reconstructed internally. In addition, differing degrees of reconstructability of sound changes must be considered. A similar problem also applies to Comparative Reconstruction, where forms and sound changes may no longer be reconstructable or no statement can be made whether two reconstructed forms were in use at the same time. The Comparative Method, therefore, is also regarded as "historical but timeless" (Shevelov 1964:5). In practice, most determinable sound alternations are of low time depth. The tendency to regularize alternations contributes to an increasing decrease in older sound alternations over the course of language history, with a naturally associated increase in new alternations. As a result, the temporal variance between reconstructions may be less pronounced than critics have assumed. The assumption of "FAR-REACHING RESULTS" (Hoenigswald 1944:78) cannot be confirmed. Only in the case of Kurylowicz's universalistic reconstructions (see Sect. 4.3.5.2), one deals with levels that are no longer temporally tangible or assessable, making the term "pre-language" undefinable. This leads to the equally common criticism that the results of IR are "MORE RECENT" than those of the Comparative Method (Chafe 1959:495) and therefore do not allow deep insight into older language stages.

This temporal vagueness also results in a lower historical significance of the reconstructions. Earlier supporters of IR still tried to grasp a rough temporal range; for instance, Pisani (1938:33) made the criticism that "la ricostruzione interna" makes us think Latin *arx* 'citadel' must be an inherited word because it belongs to an unproductive inflection in Latin. However, the lack of correspondences in other Indo-European languages would serve as an argument against the word being old. IR can infer nothing more than a relative chronology from productivity or unproductivity (see Sect. 3.5). A problem may arise, however, when the different dimensions of internal and comparative reconstruction are combined, as is done in approaches with COMBINED INTERNAL AND COMPARATIVE METHODS. Subsequently, there is a danger of using IR to reconstruct forms that are older than the proto-language of the Comparative Method.

6.3.2 The Reconstructions Give an Incomplete Picture

For all reconstruction methods, it is true that only historical elements that have left relicts can be reconstructed. Unlike for the Comparative Method, the regularization tendencies of a language necessarily entail a significantly reduced reconstruction potential for IR: IR lacks "in the quantity of phenomena taken into consideration" (Prosdocimi 1977:95). Sound changes can only be reconstructed using IR if the corresponding alternations are salient in the language's grammar. Merger is a large group of sound changes and one of the phenomena that cannot be reconstructed by IR or can only be reconstructed to a limited extent (cf. Hoenigswald 1965:132). At the supraphonetic level, lexeme substitutions ("lexeme replacement" such as the Anglo-Saxon *sinwealt* by the Middle English *round*) and syntactic or grammatical-categorical mergers are included. The criticism of an incomplete reconstruction is therefore justified.

Although other reconstruction methods cannot provide a complete picture either — note that the wordform *sinwealt* cannot be reconstructed comparatively with modern Germanic languages — the severe limitation in data also leads to a limitation in potential applications. This leads to the fact that IR can reconstruct fewer sound changes (cf. Lass 1997:241) and can only represent complex sound changes in a simplified way. Especially for agglutinative languages, which offer little alternation, only a few sound changes can be determined. For Finnish, Anttila (1973:346) was able to determine some consonantal changes, but most vowel changes remained unidentified. Milewski, therefore, came to the following conclusion in his introduction to historical linguistics:

> *The method of internal reconstruction (relative chronology) generally is of secondary significance. It enables us to determine a small number of unconnected facts, but does not enable us to reconstruct the entire evolution of a language.*[56]

[56] Milewski (1973:110)

In general, complex factors cannot be reconstructed by IR alone. This includes the establishment of controversial relative chronologies, for which IR also often yields only unsatisfactory results (see Birnbaum 1970:109). In fact, IR assumes a linear relative chronology (*linear order hypothesis*) and ignores cases of parallel or iterative occurrences of sound changes (see Chen 1976:211–213 for examples). IR often fails to fulfill its purpose of resolving controversies that have arisen from the Comparative Method. However, this assessment is not universally shared and is estimated differently in individual cases. For instance, Campbell and Grondona (2007:24) emphasized that the palatalization in Chulupí "is only suggested in the evidence for internal reconstruction but is not adequately confirmed on the basis of internal evidence alone."

6.3.3 The Reconstructions Are Speculative and Unreliable

The results of IR are often assessed by critics as "less reliable" (Ringe 2003:244), "unsafe," or even as "glottogonic speculations" (Hoenigswald 1944:78). The basis for this view is the multiplicity of reconstruction possibilities, which increasingly occur both in the reconstruction method itself and in inferring methods. IR often does not offer a clear answer for determining the direction or condition of a sound change. Instead, it usually offers only a variety of alternatives. Hock (2013:26) categorized the types of IR and assumed a gradual decrease of "reliable" results, starting with automatic, recurrent alternations, over morphologically conditioned alternations to hypotheses based on observations from typology, and universal research.

There is often a lack of comparison with other languages that could support or discard an approach. Therefore, the question of whether the procedure was carried out purely objectively or was guided by subjective judgments may arise for the linguist during the process of IR. Even supporters of IR are often unclear as to whether their result was indeed arrived at purely objectively without taking external facts into account. Fowkes correctly attributed the Welsh initial consonant mutation to a *sandhi* phenomenon, but he (1950:144) himself stated that it is "difficult to say whether synchronic evidence would actually lead to the realization that a *lost* final vowel could be responsible for the process [original emphasis, A.B.]." The question whether he could unconsciously have been looked for indications of *sandhi* rules justifiably arises. Fowkes himself also indicated that the historical cause of consonant mutation was not discovered before the emergence of comparative linguistics.

Thus, internal reconstructions remain speculative, as these reconstructions cannot be further internally tested. According to Lass (1975:18), there is only a "testability by accident." He did not elaborate on the notion of "testability" for reconstruction methods, but he likely agreed with Ringe in this respect:

> *[Comparative Method] circumvents the effects of these [for IR problematic sounds, A.B.] changes by adducing evidence from related languages or dialects in which the same changes have not occurred; IR has no comparably straightforward means of "undoing" the changes.*[57]

Consequently, assuming a sound change inferred comparatively from two languages, the cognates of third languages can confirm or reject a reconstruction. In fact, internal cognates can show analogical adaptions or foreign influences in a similar way. Internal cognates, however, exhibit higher levelling pressure.

In practice, testability is limited to results of other reconstruction methods. Birnbaum (1977:14–15) proposed a test via the Philological Method (i.e., a comparison of the result of an IR with the historical attested pre-stage, such as Latin or ancient Greek). "Method-internal" tests have not been presented so far. One might use Birnbaum's (1977:15–16) test procedure, which reconstructs two related languages internally and compares the results. However, one is unlikely two expect true "falsification" with this procedure. The absence of alternation in the second language does not say much about the internal reconstruction nor about the question of any pre-sound of a sound change in the first language. Faulty reconstructions would only be reproduced here.

6.4 Are the Conclusions Justified? Criticism of Extrapolation

Especially those approaches that work with universals or "usual rules" must justify their extrapolative procedures (cf. Anttila 1973:350). Extrapolations are estimations of behaviors beyond assured values. This is true for mathematical extrapolations,

[57] Ringe (2003:244)

making statements about a larger or different (uncertain) sample from a smaller (assured) sample. The question for IR is to what extent conclusions about other languages can be drawn from the "rules." The answer ultimately depends on the individual rule and probably on the typological similarity of the languages in question. Empirical data on the frequency of "usual rules" can be a guide as to the probability of reconstruction. However, since empirical studies of sound change are usually limited to only a few families or linguistic areas, their predictive value must be taken with caution. Anttila, who endorsed universal approaches in IR, argued as follows:

> *[B]ut linguists at least must use it in the sense that they use their background and training in attacking a new language. I.e. they expect that every new language behaves like a language. Further, it is difficult to know where interpolation, the filling out of intermediate terms in a series, ends and extrapolation begins. […] In short, analogy and extrapolation are not supposed to prove anything. They give us ideas to be tested, hypotheses, and these may ultimately lead to convincing analyses.*[58]

6.5 Conclusion

In the previous chapters, the advantages of non-comparative reconstruction have already been highlighted. Additionally, the disadvantages of IR have also been analyzed. It is not uncommon for criticisms to be generally inherent to all reconstruction methods, which has become all the more clear when the criticisms of IR are compared with those of the Comparative Method. Milewski (1973:110–111) lists as criticisms of the *Comparative* Method:

- It is uncertain whether the reconstructed elements actually existed simultaneously (i.e., there is the possibility of anachronisms)
- The incompleteness of the proto-language
- The possibility of (early) borrowing cannot be excluded
- Parallel developments in the daughter languages

[58] Anttila (1973:350)

- Typologically, the method is rather applicable only to morphological and inflectional systems

The similarities between the criticisms are noteworthy; nevertheless, IR often has far less material available to provide a detailed picture of historical linguistic development. The goal of the next chapters will be to re-evaluate the methodology of IR by means of automation, for the purpose of ultimately evaluating and optimizing proposed approaches.

7. How to Automate Internal Reconstruction

The second part of this thesis deals with the automation of internal reconstruction methods. This issue is part of the so-called *Computational Historical Linguistics* (CHL) — a branch of historical linguistics that tries to solve historical-linguistic problems using computational approaches. In addition to the reconstruction of proto-languages, this also includes the reconstruction of cultural issues, such as the locality determination of ancestral homelands (Boukaert et al. 2012 for Proto-Indo-European, Sicoli and Holton 2014 for Dene-Yeniseian) and matrilocality (Jordan et al. 2009), or the study of cross-linguistic features (e.g., de Boer 2000). In doing so, CHL relies on methods from other disciplines, especially computational linguistics and bioinformatics. An overview of the methods, goals, and application areas of CHL is provided by Jäger (2019); for a rough overview of existing problems and questions, I refer to List (2019). While some linguists are skeptical of quantitative methods in historical linguistics, the last two decades have shown an increasing interest in CHL by classical historical linguists (e.g., Hock/Joseph 2019:472, Garrett 2015:241). At the same time, CHL moved away from traditional methods of lexicostatistics, which are associated with several problems, such as the incorrect or inaccurate translations of basic vocabulary in individual languages or the erroneous identification of loanwords. Geisler and List (2009:1) tested recent stochastic methods based on lexicostatistical datasets and concluded that the problems "are still so grave that no reliable results can be drawn from" lexicostatistical methods. Phylogenetic methods, on the other hand, are shown to be much more robust against loanwords, as demonstrated by Greenhill et al. (2009).

Implementing IR computationally poses some fundamental questions, starting with an appropriate definition. This thesis follows the synchronic-monolingual definition of IR with a focus on internal means (see Sect. 3.1). The use of external means (e.g., universals or other cross-linguistic knowledge) offers great possibilities for future works on the topic but will remain contestable due to their extrapolative nature. With the move away from the structuralist definition, automated IR is given more opportunity. In fact, many languages show little irregularity in paradigms, so automating the paradigmatic morphophonemic method would add little scientific value to many languages. Moreover, in general, quantitative methods are of particular interest for poorly studied languages and proto-languages; but it is precisely true for these languages that the morphophonemic method is often of little use. For example, the linguist decides whether an alternation is proto-lingual or not and how it worked in the proto-language. This implies pre-linguistic considerations of the linguist which are merely reproduced by the computational reconstruction. Other implementation issues are discussed in more detail in this chapter.

7.1 Related Works

7.1.1 Internal Reconstruction and Computational Historical Linguistics

The topic of "internal reconstruction" has, so far, received little attention in CHL. For the most part, the specific literature is limited to a mention of it in footnotes (e.g., Moran et al. 2020:87 fn. 15). According to List (2018:30), there are two reasons to note for this: Inflectional paradigms vary widely across languages, making it difficult "to come up with a common way to investigate them," and secondly, irregularities are the exceptions, so one inevitably encounters a sparse-data problem.

The advantages of INCLUDING IR IN AN AUTOMATED COMPARATIVE RECONSTRUCTION PROCEDURE were nevertheless seen by different linguists. Hewson, an early proponent of CHL who took a computer-aided approach, saw the possibility of using semi-automatic IR to identify undiscovered word formatives in proto-languages. In doing so, "the data is organized into a concordance [sorted by internal word formatives] by a simple computer program [… and] the linguist is enabled to survey a remarkable range of immediately relevant data" (Hewson 1989:158). Koskenniemi (2013:51), whose approach models historical relations between languages using finite-state transducers, saw the use of morphophonemes rather than phonemes as an advantage in mapping cognate languages. These morphophonemes could get through IR. Adelaar (2013:123), on the other hand, called for a determination of etymons using IR before comparative comparison in order to reduce problems of lexicostatistics. However, the IR process has not been automated by any of these three linguists.

IR also plays a more frequent role in quantitative linguistics when considering the RECONSTRUCTION OF SEMANTICS. The theoretical linguist Starostin (2016:184) described a method for reconstructing proto-word lists from Swadesh lists of successor languages and considered techniques of IR for determining the archaic proto-form for a concept when daughter languages provide multiple etymons (so-called "onomasiological reconstruction"[59]). For this issue, he defines so-called "distributional" and "extra-distributional rules," which are largely inferring methods by comparing the data in the daughter languages (*distributional rules*) or by incorporating additional knowledge (*extra-distributional rules*). Extra-distributional rules also include inferring methods familiar from the grammatical IR of Sect. 4.3, for example, the assumption that a word is younger if "its derived character can be demonstrated either on the synchronic level of language A or on the diachronic level of the protolanguage" (2016:186). However, CHL approaches to onomasiological reconstruction that considered Starostin's rules have so far been limited to non-extra-distributional rules only (e.g., Kassian et al. 2015:307, Jäger and List 2018:39).

List's (2018) proposal of implementing IR is based on the semantic method and is meant to reconstruct meanings from compounds. As an example, he chose the word $tɕ^{h}i^{55}$ (the high number represents a tone), which occur in the Burmese language Achang as a word for 'arm', 'claw', 'foot', 'hoof', 'knee', 'leg', 'thigh', and 'ant'. He formalized the method with a bipartite network, which is a network with two kinds of nodes, that are used to represent different types (cf. Hill and List 2017). In this case, one type formed the wordforms, the second the meanings "attributed to the sub-parts of the words" (List 2018:32). The method itself has not been implemented and the example remained anecdotal (cf. List 2018:33).

7.1.2 Comparative Method and Computational Historical Linguistics

7.1.2.1 Ancestral State Reconstruction

A much-discussed subfield of CHL is Ancestral State Reconstruction (ASR), in which parts of the Comparative Method for the machine reconstruction of proto-forms are automated (for an overview, see Jäger 2019:170–175). The sub-processes include the identification of genetic relationships, the detection of cognates as well as loanwords, or the verification of the validity of sound laws. A representation of a pipeline with all subprocesses of the automatic Comparative Method is provided by Steiner et al. (2011). Each of the subprocesses has special preconditions

[59] In contrast to "semantic reconstruction," which is concerned with reconstructing the original meaning of a word, "onomasiological reconstruction" tries to reconstruct the word for a certain concept in the proto-language (cf. Jäger and List 2018:24).

and problems and requires its own method, which makes them their own subfield of CHL. A complete implementation of all steps has not yet been realized.

Attempts to automate the Comparative Method are surprisingly old. Kay (1964) has already implemented the first program for an IBM 7090 computer at the time of magnetic-core memories. For this, he used a RULE-BASED algorithm with truth-functional expressions, but it was hardly suitable for the technical capacities (cf. Kay 1964:V). From four English-German cognates (*on-an*, *nut-Nuß*, *that-daß*, *bath-Bad*), he was able to derive some sound correspondences, but the author estimated “that it would take some four or five hours of computer time to analyze a list of hundred pairs of forms” (1964:18). Hewson (1974) tried to implement the Comparative Method by machine, but in doing so, he restricted himself exclusively to cognate detection in languages whose sound laws and relationships are already taken as given (cf. Hewson 1974:194–196). Nevertheless, with his *Computer-Generated Dictionary of Proto-Algonquian* (Hewson 1993), he succeeded in producing the first such work with over 4,000 reconstructed wordforms from four Algonquian languages (Cree, Fox, Menomini, and Ojibwa). Rule-based approaches were also tried in later times (e.g., Oakes 2000) but proved to be non-robust in dealing with exceptions in the form of analogy or borrowing. Then, PROBABILISTIC APPROACHES followed as the second generation of implementations of the Comparative Method. One of the advantages of these methods is that they are able to deal with exceptions appropriately. The most recent phase of CHL is decisively shaped by computational biologists, such as Russell Gray, Robert McMahon, or Alexandre Bouchard-Côté, and uses PHYLOGENETIC APPROACHES (Gray and Atkinson 2003, McMahon and McMahon 2005, Pagel et al. 2005). In this phase, the method of Bayesian phylogenetic inference — originally developed for the evolution of gene sequences — is adapted to the phylogenetic tree model of historical linguistics. Bouchard-Côté et al. (2013) presented a purely probabilistic framework for reconstructing ancient wordforms, which they evaluated with Austronesian cognates. A comparison of the machine-produced reconstruction with a manually produced gold standard yielded 85% agreement, and about twice as much disagreement as is common between manual reconstructions by two different linguists (cf. 2013:4224 and 4226). Despite this successful result, the automated Comparative Method is still in its infancy. Austronesian has relatively simple phonology and the proto-language is younger compared to other language families. A critical stance on applying Bayesian phylogenetic approaches to historical linguistics is taken by Pereltsvaig and Lewis (2015).

7.1.2.2 Reconstruction of Sound Change

The models of Bouchard-Côté et al. (2013) and many other computational historical linguists are designed for proto-form reconstruction (Ancestral State Reconstruction). Reconstructing sound changes in these models proves to be quite difficult due to the modeling of context in such a way that they can represent sound correspondences without exception (cf. Dellert 2019:38–39). The framework of Bouchard-Côté et al. (2013:4228) takes the previous sound into account and is thus not able to deal with metatheses. Sound changes play a central role in the model of Hruschka et al. (2015), in which sound change is equated with concerted evolution in biology. Their statistical model detects concerted changes in cognate data of Turkish languages and, in addition to sound change, also inferred phylogenetic trees indicating when and at which node a sound change is to be applied. Although the procedure can also be carried out with phonetic context (Hruschka et al. 2015:8), the authors restricted themselves to sound correspondences without context (likewise Wheeler and Whiteley 2015 did for Uto Aztecan). Accounting for context imposes a significant cost on statistical models, and correctly determining context requires a special degree of *fineness* that models struggle to achieve. Quantitative models describing a phonetic change as abstract sound laws with distinctive features (e.g., C[+voiced] > C[-voiced] / _C_0[-voiced]) are lacking to date.

7.1.3 Phylogenetic Methods and Internal Reconstruction

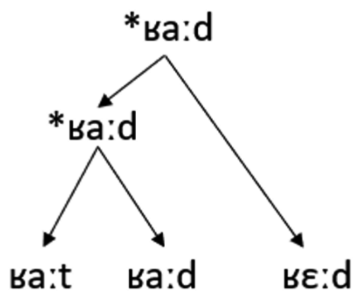

Figure 7.1: Sample of a phylogenetic tree with wordforms of the German word *Rad* ‘wheel’.

Papers that use phylogenetic methods for IR have not yet been published. The obvious reason is the lack of a phylogenetic model. Phylogenetic trees in IR are, however, brought into play in the field of the relative chronology of sound changes (e.g., Anttila 1973:323). If we allow a relative chronology of wordforms for purely theoretical considerations, we end up with trees as illustrated in Fig. 7.1. Since, by definition, IR cannot include external languages, the allomorphs must be defined as (internal) word correspondences for a phylogenetic approach. In this case, paradigmatic categories and derivations take the place of external languages, as seen in Tab. 7.A. With exception of suppletion, only cognates are in such a line, so that a step of cognate clustering (cf. List et al. 2017, Dellert 2019:124–137) plays a subordinate role here. One of the main problems of this approach is the irregular distribution of the phonetic alternations in the grammatical categories. In the Comparative Method, a sound change in a column is ideally regular. In the IR, however, one and the same sound change (e.g., the *umlaut* in *R**ä**d-er* and *z**ü**ng-eln*) is distributed over several columns and is not completely continuous for one column. This makes the detection of a sound change more difficult and becomes clear when we add another final devoiced form of *ʁɛːd* to the example in Tab. 7.A.

Table 7.A. Sample of ‘internal word lists’ for German. Grammatical categories replace the languages of cognate lists used for the Comparative Method.

Concept	**Nom. Sg.**	**Gen. Sg.**	**Nom. Pl.**	**Verb with *-eln***
wheel	ʁaːt *Rad*	ʁaːd *Rad-es*	ʁɛːd *Räd-er*	ʁaːd *radeln*
dog	hʊnt *Hund*	hʊnd *Hund-es*	hʊnd *Hund-e*	-
tongue	tsʊŋə *Zunge*	tsʊŋə *Zunge*	tsʊŋə *Zunge-n*	tsʏŋ *züngeln*

This would infer two possible phylogenetic trees, as illustrated in Fig. 7.2, which no longer provide an unambiguous reconstruction. A sound change, such as final devoicing, applies across all word forms, which would create numerous new leaves in an IR-phylogenetic tree. Alternatively, the sound changes themselves could be defined as “languages” (e.g., *umlaut*: [ʁɛːd], [t͜sʏŋ] and *final devoicing*: [ʁaːt], [hʊnt]). But then, the linguist would give important information to the machine that it should automatically detect. Despite these arguments, phylogenetic automation of IR seems possible with some adjustments. However, a phylogenetic approach hardly seems to provide a decisive advantage compared to rule-based or statistical approaches. This is especially true for automations that aim at reconstructing sound change. The focus of the approach in this thesis will therefore be on non-phylogenetic methods.

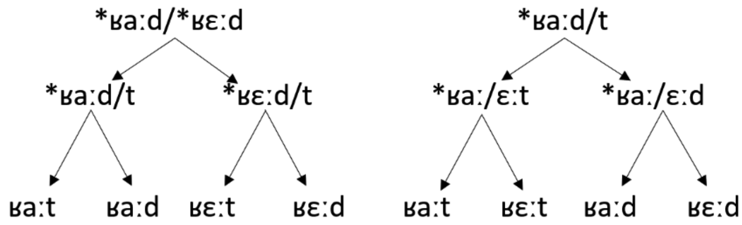

Figure 7.2: Two possible phylogenetic trees for the wordforms of *Rad* (cf. Tab. 7.A). The best tree must be derived by a subsequent procedure.

7.2 Data for Automating Internal Reconstruction

The prerequisite for computer-assisted studies is digitally available data sets. Depending on the aim of the study, these data must be available in an appropriately structured form (e.g., as corpora or lexical databases). CHL mostly works with lexical databases, which ideally provide the vocabulary of different languages in transcribed form. The lack of such databases is an obstacle to computer-assisted historical linguistic research, but databases of various types have already been created and have reached a useful size. One type of them contains *basic concepts* collected for a variety of languages. The largest of these databases, with currently over 400,000 synsets, is the *Automated Similarity Judgment Program* (ASJP, Wichmann et al. 2020), which has cumulated words for 40 basic concepts from over 9,700 languages and language varieties to date. Comparable databases have over 1,000 different concepts but are limited to one language area, such as TransNewGuinea.org (Greenhill 2015) or NorthEuraLex (Dellert et al. 2020).

The second type of database is the *cognate database*, which usually do not provide wordforms but information about whether a language contains a cognate form. Their creation requires much more technical and research effort. The largest databases of this type are IELex (Dunn 2012) for Indo-European, UraLex (Syrjänen et al. 2013) for Uralic, and the Austronesian Basic Vocabulary Database (ABVD, Greenhill et al. 2008). Each of them contains little more than 200 concepts.

For automation of IR, the word lists of the individual languages in these databases are hardly suitable. The few concepts and the lack of allomorphs is a limiting factor for automated IR, which may explain the lack of comparable studies in CHL.

The morphophonemic method requires a database of allomorphs, which is not available in comparable dimensions. The required morphological data can be extracted from electronically available dictionaries. Alternatively, word lists (with paradigmatic wordforms) can be created, and their words can be decomposed into single morphemes using a morphological analyzer (for an overview, see Carstensen et al. 2010:236–263). However, these possibilities are not available for less intensively researched languages. This eliminates an area of application for which automated reconstruction would generally be of interest.

7.3 Structure of the Thesis

Chapter 4 presented a number of methods that are classified as IR in the literature. Many of these are not suitable for automation for various reasons. Grammatical reconstruction methods show, on the one hand, the problem that the required data are not necessarily available in digital form. This applies, for example, to the archaistic method, which requires, among other things, information about “archaistic” connotations, or Kurylowicz’s method, which works with “markedness” and “primary/secondary functions.” Although some dictionaries provide the information “obsolete,” this information is more diachronic and less connotative. On the other hand, methods such as Kurylowicz’s or Bauer’s algorithm are extremely elaborate and highly individual. In both cases, features are always included as a decision factor, which seems to have a certain degree of arbitrariness. Kurylowicz brings into play the lexical stress whenever it is needed and Bauer’s algorithm lacks an approach for determining the main characteristics of a language. The thesis will therefore be limited to PHONOLOGICAL RECONSTRUCTION METHODS. Of these, chapter 4 discussed five different methods:

- Morphophonemic method
- Distributional method
- Pattern reconstruction
- (Phonological) archaistic method
- Typological and onomatopoetic method

While the phonological archaistic method shows a problem with lack of the required information, the situation is more complex with pattern reconstruction. The method according to Latta shows parallelism to the morphophonemic method and may in many cases merely represent an alternative variant of it (see Sect. 4.2.3.2). Since it anticipates the process of abstraction, it is the more procedurally complex variant. The reason for this is that the patterns must be defined at the outset. This would result in a large number of possible patterns for a morpheme such as *min*: *nasal-vowel-nasal*, *labial-front_vowel-nasal*, etc. These patterns all have to be integrated into the process and noticeably affect the processing time. The typological and onomatopoetic methods can be summarized from a procedural point of view. Both have in common that they start from pan-lingual knowledge, which must first be determined and finally integrated into the reconstruction process.

In this thesis, six different phonological sub-methods are automated. Their implementation and the purpose of automation vary considerably. While the procedure of the morphophonemic method has already been formulated in detail since Hoenigswald (1944), the implementation of the distributional method has yet to be concretized. For the distributional method, this is connected with the question of the feasibility and significance of the method in principle, which has already been discussed in detail in the IR literature. The six approaches carried out here can be assigned to two phonological methods from the literature (see Fig. 7.3). These approaches will be discussed in detail in the next chapters. Their algorithms used for automation will be presented, evaluated, and the results will be discussed with respect to the objectives of this thesis. In the last chapter, a summary follows, in which the results of the different methods are compared.

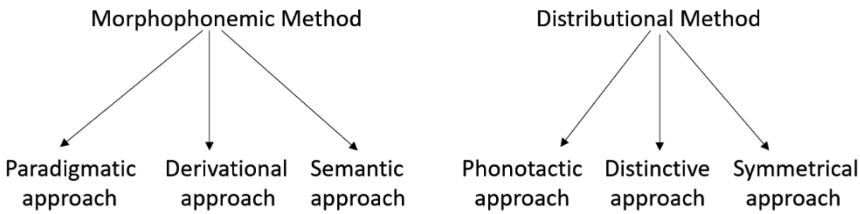

Figure 7.3: Overview of the five IR approaches picked out as central themes in this thesis. These can be assigned to morphophonemic and distributional methods.

7.4 Aims of the Thesis

The first aim of the work is to AUTOMATE THE PROCESS OF IR for the first time, that is to say, to reproduce the process to such an extent that it can be processed computationally. Unlike for the Comparative Method, which is largely uniformly defined, some IR methods are insufficiently formulated. For the distributional method, this also includes an algorithmic definition, while the focus of the morphophonemic methods is rather on the way of implementation (i.e., the choice of data and their integration). Another goal is also the experimental TESTING OF NEW AND INNOVATIVE APPROACHES. By integrating larger amounts of data, future IR has many more possibilities than a simple analysis of alternations. While automated Comparative Method aims, among others, to explore macrofamilies, I see the purpose of an automated IR as discovering new and optimized methods, which in turn will eventually provide better insight into the antecedents of isolated or proto-languages. A further objective is to EVALUATE THE DIFFERENT METHODS and to address the general question: What can IR reconstruct in purely objective terms — both in general and, more specifically, with automated methods? Furthermore, it should be possible to answer the following concrete questions:

- How old are the internally reconstructed sound changes?
- Is it rather phonological rules or sound changes that are reconstructed?
- Which resources are suitable for automated IR?

No aim of this thesis is to establish sound changes newly detected by these automated methods. Since "the null hypothesis in historical linguistics is to deem words to be unrelated unless proved otherwise" (Bouchard-Côté et al. 2013:4226), all identified sound changes are considered false if they have not already been proclaimed in the literature.

7.5 Evaluation

7.5.1 Evaluation Measures

In this section, the question of how to evaluate the results is explored. Despite the lack of evaluation standards in CHL, standard methods from general research can be used in most cases. The standard measures for binary classification tasks consist of ***recall*** and ***precision***, which compares the results of the automation process with a gold standard. Precision is calculated from the number of positive objects (in this case, *positives* are "found" sound pairs) that are correct (i.e., also occur in the gold standard) divided by the number of sound changes found in total. The following $-k$ example contains six correct out of twelve sound pairs:

$$precision = \frac{r_p}{r_p + f_p} = \frac{6}{6+6} = 0.5$$

Recall gives the value of how many pairs that belong to the sound changes of the gold standard were recognized. This is calculated by dividing the number of correctly found pairs by the number of pairs in the gold standard (the example has eight gold standard pairs).

$$recall = \frac{r_p}{r_p + f_n} = \frac{6}{6+2} = 0.666667$$

As a measure that combines both precision and recall, there is the F-score, which is calculated as follows:

$$F = 2 \times \frac{P \cdot R}{P + R} = 2 \times \frac{0.5 \ \cdot 0.6667}{0.5 + 0.6667} = 0.571429$$

Evaluating the results proves to be difficult because no gold standard exists on manually produced internal reconstructions. Papers attempting to evaluate IR (such as Fowkes 1950 or Anttila 1973) performed IR for concrete languages and compared their results with those of the Comparative or Philological Methods. The comparison is lame for two reasons: First, internally reconstructed sound changes are not identical to comparative reconstructed sound changes and both methods reconstruct different language stages (proto-language vs. pre-language). The conditions of the German *umlaut* cannot be reconstructed internally but the reconstruction of sound correspondences is possible. Second, IR "reconstructs" not only sound changes but also synchronic alternations, which do not occur in the comparative gold standard data. For evaluation purposes, therefore, it makes sense to define both sound change and phonological rules as positive targets but to consider them with different gold standards. From the different reconstruction stages, the question of the point of comparison arises: do sound changes from Indo-European times still belong to the gold standard? These questions ensured that a quantitative evaluation of IR has been lacking so far. Another question is when an identified sound change is considered "correct." Is the correct sound correspondence (sound pairing) sufficient for this or must the condition also be correctly determined? Since the condition can often only be determined inaccurately with IR, the latter evaluation would lead to an extremely high error rate. In this work, therefore, sound correspondence and conditions are evaluated separately. The third evaluation is the direction of the sound change because inferring methods are used for this.

7.5.1.1 Match Rate

To answer the question of what IR as a reconstruction method can generally reconstruct, a MATCH RATE is determined. It indicates how many of the sound changes found by the automated process can actually be assigned to a sound change or phonological rule from the gold standard without fulfilling a strict condition. In a simplified way, it is assumed that the method can reconstruct a sound change in principle, but the exact determination of conditions or the direction of the sound change can be attributed to the optimization needs of the automation. The match rate is therefore mainly used for estimating the method, less for evaluating the concrete automation. The rate is relatively easy to meet, and a detected sound change may correspond to several data from the gold standard. The measure is therefore also used to detect "clear" errors.

7.5.1.2 Evaluation of the Sound Correspondences

Table 7.B. Sample of an evaluation of an abstract sound change "voiceless-voiced". The sound change of the gold standard is *[-sonorant] > [-voice] / _C_0[-sonorant]*. The data are taken from paradigmatic method with German morphemes. A missing match (false negative) is in brackets.

Sound Change "Voiceless-Voiced"	Value	Gold Standard
g-k	0.692642	g > k
p-b	0.679333	b > p
z-s	0.623308	z > s
g-ç	0.561831	(ʒ > ʃ)
t-d	0.559366	d > t
ŋ-ç	0.413951	
v-f	0.402418	v > f
χ-ŋ	0.346873	
ŋ-x	0.289705	
ʀ-s	0.119305	
g-f	0.075628	
m-f	0.075312	
p-m̩	0.083336	
χ-g	0.082840	
t-n	0.070811	
t͡ʃ-d͡ʒ	0.0	d͡ʒ > t͡ʃ
ʔ-b, ʔ-l, t͡s-d, ʀ-t, v-h, t-b, ʔ-ʀ, ʃ-g	0.0	

For the evaluation of the sound correspondences, the precision of the correct sound pairs of all matched sound changes is calculated. The evaluation of abstract sound laws is more complicated. For these, the sound changes of the gold standard are decomposed into single sound changes. True positives are all single alternations which are both in the set of the detected sound change and the set of the corresponding sound change or phonological rule of the gold standard. As an example, the detected abstract correspondence *voiceless:voiced* matches the phonological rule [-sonorant] > [-voice] / _C_0[-sonorant] (see Tab. 7.B). In this way, the evaluation measures recall and precision can be calculated. Precision is calculated from the number of "found" sound pairs that are correct (i.e., the pairs with counterparts in the column "gold standard" of Tab. 7.B), divided by the number of all sound changes.

$$precision = \frac{r_p}{r_p + f_p} = \frac{6}{6 + 18} = 0.25$$

The calculated precision for the example of 0.25 is low, considering that the feature of sound change (*voiced*) was correctly determined. This has to do with the fact that this sound change requires more than one feature pair *voiced:voiceless*. The sounds must be a sonorant of the same manner and place of articulation. Among the values of more than 0.5, the precision is significantly higher than 0.8. The only false positive more than 0.5 is the pair *g-ç*, which represents another sound change. The recall is calculated by dividing the number of pairs without brackets in the right column of Tab. 7.B by the total number of pairs in that column:

$$recall = \frac{r_p}{r_p + f_n} = \frac{6}{6 + 1} = 0.857142$$

The F-score, which combines precision (P) and recall (R), is calculated as follows:

$$F = 2 \times \frac{P \cdot R}{P + R} = 2 \times \frac{0.25 \cdot 0.85714}{0.25 + 0.85714} = 0.3870974$$

7.5.1.3 Evaluation of the Change Direction

As discussed in the first part of the thesis, the internal reconstruction method itself does not give any way to determine the pre-phoneme. Only inferring methods proposed by the research give an indication of the sound change direction. The evaluation of direction in this thesis aims to evaluate and compare three of these methods:

- phonetic plausibility,
- the articulatory similarity between sound pair and conditions and
- rule of restricted sounds.

The number of correct sound change directions divided by the number of matched sound changes gives the PRECISION OF THE CHANGE DIRECTIONS.

7.5.1.4 Evaluation of Conditions

Table 7.C. Sample of an evaluation of conditions. The data are taken from paradigmatic method with German morphemes. False conditions are in brackets. The threshold was set to 0.95.

	Detected Sound Change *Near Open → Uvular*	**ʁ > ɐ / _C$_0$ (Adjusted to ɐ > ʁ / _V)**
0.998571	vowel|_|vowel	_V
0.998296	front|_|vowel	_V
0.998243	unrounded|_|vowel	_V
0.997175	_|vowel	_V
0.994536	_|central	_V
0.993289	(central|_)	
0.993103	_|front	_V
0.992958	_|unrounded	_V
0.987805	plosive|_|unstressed	_V
0.984127	_|closed	_V
0.980769	voiceless|_|unstressed	_V
0.972222	voiced|_|unstressed	_V
0.970588	bilabial|_|unstressed	_V
0.970588	(a_)	
0.967742	(#_)	
0.962963	alveolar|_|unstressed	_V
0.956522	#_e	_V
0.954545	plosive|_|open	_V
0.952381	plosive|_|rounded	_V
0.950000	plosive|_|close-mid	_V
0.950000	velar|_|unstressed	_V
0.95	(ɔ_)	

A different procedure is used to evaluate the conditions, since different conditions may be valid. For instance, the German change *r* > *ɐ* applies to the conditions _C and _#, so automation may need to capture both conditions as correct. Therefore, a threshold value of at least 0.9 is introduced: Any condition above this value is considered positive (i.e., found) in the evaluation. To get older sound change conditions, a threshold lower than 0.9 could be used, but this could lead to a wide variety of irrelevant conditions. Since the sound change direction is already evaluated differently, the conditions and sound pairs of the gold standard are adjusted so that an incorrect sound change direction is not included in this measure. In the example of Tab. 7.C, the threshold for the feature pair *near open → uvular* was set to 0.95. The correct condition of this sound change is _V, which appears in fourth place with a value of 0.9972. A large number of the positives are duplications of this condition, which are difficult to avoid. These positives are regarded as "correct" because they are variants of the gold standard's condition. This gives a precision of 0.846 but with a duplication rate of 95.46%. The duplication rate can be lowered by a higher threshold. However, it should be noted that this could reduce the precision for diachronic sound changes.

7.5.2 Evaluated Languages

In this thesis, the methods are evaluated on two languages with different gold standards. GERMAN is the main object of investigation and provides a gold standard of sound changes that have been made accessible by the comparative, external, and philological methods. German has been documented since the 8th century and is divided for the evaluation into the stages *Old High German* (OHG)*, Middle High German* (MHG, from the 11th century), and *Early and New High German* (NHG, from the 14th century). From this, three different gold standards can be inferred, representing different time depths: sound changes from Early Germanic to OHG, from OHG to MHG, and MHG to modern German. The fourth gold standard contains exclusively phonological rules of German. It is expected that the precision values will decrease with older language stages. The gold standard for diachronic sound changes is based on Kümmel (2007) and takes into account all changes listed there as "…hd" (i.e., … High German) and "g-…hd" (i.e., common … High German) as well as dialectal if the sound change was adopted in today's standard German. The vocalic sound changes originate from Paul et al. (1969) and had to be put into rule form for this purpose. A complete list of all considered sound changes can be found in the appendix. The synchronic phonological rules were taken from O'Brien and Fagan (2016).

PROTO-INDO-EUROPEAN (PIE) was used as the second evaluation language. As a proto-language, all reconstructions of Pre-Proto-Indo-European have been reconstructed internally by linguists. Thus, this evaluation primarily serves to compare the automated IR with a "manually" internally reconstructed gold standard. Proto-Indo-European is dated to about 5,000–3,000 BC (Meier-Brügger 2010:194) and is the best-researched proto-language. However, there is no survey of Pre-Proto-Indo-European sound changes that are generally accepted by most Indo-Europeanists. For the evaluation of phonological rules, the rules from the corresponding chapter in Fortson (2010:69–72) have been chosen, which, however, has by no means any claim to completeness.

The situation is more difficult with diachronic sound changes since there is no *opinio communis* in research. The question, therefore, is rather whether a sound change has been proposed in research as a diachronic sound change. The quantitative evaluation must give way here to the qualitative discussion. Instead of recall, a loose comparison with the internal reconstructions is proposed in order to get a rough estimate of how many sound changes were not recorded. It is also expected that the reconstructed changes depend on the source data. Only sound changes or phonological rules are reflected in the reconstructions that are in accordance with the editors' opinion. This circumstance additionally complicates a comparison of the automated results with a manual gold standard.

8. Morphophonemic Method

The morphophonemic method represents the traditional method of IR and, as such, has already been presented in detail in the first part of the thesis. Its starting point is the hypothesis of an etymological relationship of allomorphs. In the theoretical part, it was shown that this "etymological allomorphy" can be generated by means of three types and that IR is traditionally limited to the paradigmatic pairing. For the automation of the procedure, which requires information about the relation of the comparing morphemes, a separate consideration of these three types makes sense:

- PARADIGMATIC PAIRING
  Source data are wordforms of a lexeme. One would expect here an extraction of the vowel in *sing:sang:sung*.
- DERIVATIONAL PAIRING
  Source data are derivations and composites (e.g., *three:thirteen*).
- SEMANTIC PAIRING
  Source data are lexemes with the same root, but the derivation is no longer recognizable with synchronically productive word-forming means (e.g., *fox:vixen*).

The automation of these three types has different goals. The procedure and the possibilities of the paradigmatic method are well known to linguistics and can be largely transferred to the derivational method. The remaining issue for these methods, therefore, is to obtain data from digital resources, to rank the potential sound changes (this includes the evaluation and the definition of a threshold to separate "irrelevant" alternations) and to formulate abstract sound laws. Its application in this thesis will be limited to German because proto-languages are not suitable due to the lack of data. The methods would only reproduce the alternations made by historical linguists.

For the semantic method, the starting point is very different. Semantic pairing plays mostly no role in the literature. Due to the lack of synchronic means of word formation, the common origin of both allomorphs cannot be directly inferred. The pairing is based solely on a loose semantic similarity and can therefore no longer be called "morphophonemic" (but still morphological). Accordingly, due to the higher probability of random similarities a larger error rate can be assumed here. Since the literature nearly only presents this method on the basis of subjectively selected examples and has not proposed an algorithm for determining the internal cognates, the focus in Chapter 9 will be on the objective feasibility and validity of the semantic method. Altogether, this results in three variants for the morphophonemic method. Fig. 8.1 shows an overview of the individual steps of automated morphological IR: starting with a preprocessing, in which the allomorphs are determined, followed by a determination of sound correspondences and conditions, and finally a transformation of the sound pairs into sound laws.

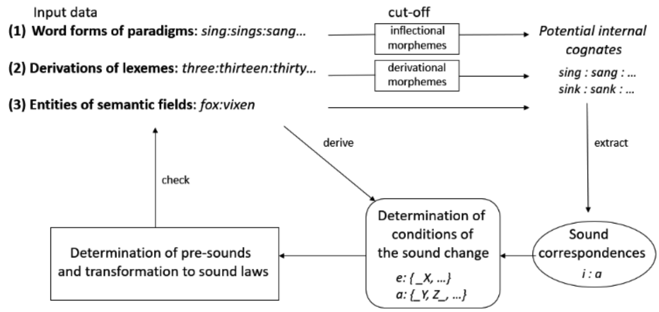

Figure 8.1: Delineation of automated morphological IR with three different input data.

8.1 Procedure of Paradigmatic and Derivational Method

8.1.1 Preprocessing

8.1.1.2 Input Data

Due to the strong methodological similarity of the paradigmatic and derivational methods, these two are treated together in this chapter. For both methods, the required data can be extracted from the German version of the online dictionary Wiktionary[60]. The data is free, large in volume, and provides both paradigms and derivations for each lemma. In addition, an IPA transcription is given for many wordforms. After sorting out unnecessary data (dialect words, river names, etc.), 92,734 lemmas remained for the paradigmatic method and 28,035 for the derivational one. With the inclusion of the online dictionary, the possibilities of automated

[60] The data of Wiktionary is available on https://dumps.wikimedia.org/. For this thesis, all data were extracted from the version of November 2018.

IR are limited on the languages available there, but the scope of available languages is larger than comparable digitally available data collections.

8.1.1.2 Removal of Inflectional and Derivational Morphemes

After the extraction of the wordforms, the word stems of the forms must be determined, the so-called *stemming*. Computational *stemmers* are tools that reduce a wordform to the stems it contains. Stemmers are freely available for several languages (with varying error rates) and are mostly based on the *Porter stemming algorithm* (Porter 1980), which was developed specifically for "suffix stripping" and is usually not applicable to IPA datasets. Other approaches of automatic morpheme detection only work well for very large datasets (Creutz and Lagus 2005). Difficulties in automatic morpheme analysis arise mainly from a lack of integration of semantic and grammatical aspects (e.g., genus). For this reason, an own-developed stemmer for German IPA data was implemented for this purpose, which drew its data from Wiktionary and should enable more linguistic stemming.

In order to be able to process compounds as well, an automated compound analysis is required. Various tools are also available for this purpose (e.g., JWordSplitter) but again, usually not for IPA transcriptions. For example, JWordSplitter shows a segmentation error rate of 11% (Abels and Hahn 2005:9), which entails a larger number of incorrect allomorphs. In this thesis, the information about compounding was taken directly from Wiktionary, which is provided in the data. This information is usually more reliable than automated analyses, but it is not found for all entries on compound nouns. Compounds are generally less important for IR due to the strong tendency of regularization, so the additional use of compound analyses is not pursued here. In addition, many relevant compounds (e.g., *woman* from *wife-man*) are no longer perceived synchronically as belonging to the corresponding lemma (*wife*) and therefore do not appear in the dictionary as a compound with *wife*.

8.1.2 Determination of Sound Correspondences

8.1.2.1 Implemented Approach

The result of the preprocessing procedure is a list of allomorphs of a word stem including all wordforms and derivations respectively. In this way, the paradigmatic variant yields four allomorphs for the German lexeme *helfen* 'to help':

- *helf → ich helf-e* (1st sg. pres.), *wir helf-en* (1st pl. pres.), …
- *hilf → du hilf-st* (2nd sg. pres.), *er hilf*-t (3rd sg. pres.)

- *half → ich half* (1st sg. pret.), *wir half-en* (1st pl. pret.), …
- *holf → ge-holf-en* (past participle)

These stems are traced to a common root according to the basic hypothesis of IR, and each reflects a potential sound change. In the next step, the sound correspondences have to be determined. The determination of sound correspondences (*Phoneme Sequence Alignment*) is an important subtask of CHL. A simple approach is to use a *Levenshtein-distance matrix* for this purpose (Levenshtein 1966). Levenshtein distance is a commonly used measure of the similarity of two-character strings through constructing a matrix and having each element of the matrix represent a pair of characters from both strings (Fig. 8.2). From the first to the last position of the matrix, all possible *traces* are traversed and for each character mismatch, the distance value is increased by one unit at the corresponding position. The *Levenshtein distance* of two strings is regularly determined by the trace with the smallest distance value.

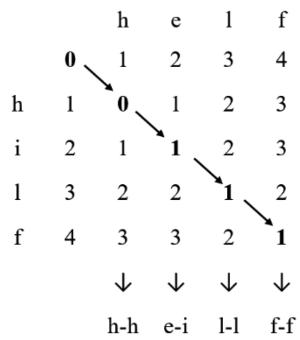

Figure 8.2: Sample of a Levenshtein-distance matrix for the German allomorphs *helf* and *hilf*. The trace with the smallest Levenshein distance is highlighted in bold.

For the example of Fig. 8.2, the Levenshtein distance is 1. By means of a back-trace procedure, the best decision path is recursively traced back and the sounds from both strings are aligned to each other (cf. Lovitt 2007). The matrix in the example thus yields the sound correspondences *h-h, e-i, l-l* and *f-f*. Besides Levenshtein distance, similar distance-based measures exist in research, such as the *longest common subsequence* (LCS, Chvátal and Sankoff 1975). It measures the longest common sequence that can — in contrast to a *substring* — be "interrupted" in the strings. The longest common sub*string* in the example would be *lf*, the longest common sub*sequence* would be *hlf*. For purpose of IR, LCS does not generate any added value.

In some cases, the back-trace algorithm cannot determine an unambiguous trace — what can happen in the case of an unequal number of sounds: for the allomorphs /*denk*/ (from *denken* 'to think) and /*dax*/ (from *gedacht* 'thought (participle)'), both the correspondences (1) *d-d, e-a, n-x, k-Ø* and (2) *d-d, e-a, n-Ø, k-x* are possible. In the procedure of this thesis, both back traces were chosen in this case. Correspondences that are incorrect should, however, not be significant in the mass of data.

Alternatively, if there are several possible correspondences, one can be selected as more probable. Intuitively, the correspondences *n-Ø* and *k-x* seem phonetically more plausible than *n-x* and *k-Ø*, so that the back trace (2) would be preferred. This additional phonetic information can be integrated into the procedure (e.g., via the articulatory similarity of the sounds), but it is much easier to integrate such different change probabilities into the procedure via a Needleman-Wunsch algorithm which obtains its information from cross-linguistic data (see Sect. 9.2).

8.1.2.2 Alternative Approaches

More complex methods have been proposed in CHL to optimize sound correspondence detection. For example, Kondrak (2002) adopts machine-translation models used to find word translations in bilingual texts and applies them to bilingual word lists to find sound correspondences. The difficulty of the task here is that the method has to distinguish cognate and non-cognate word pairs. His system "does well even the number of non-cognate pairs is more than double the number of cognate pairs" (2006:6) but solely for closely related languages.

Dellert (2019:106–113) presents the Information-Weighted Sequence Alignment (IWSA) method, which takes the information density of a sound or sequence in the wordform into account. The German verbal endings -*en* or -*t* occur more frequently in verbal forms and therefore have a lower information density than the sounds of the stem. For the verb *vergehen* 'to go by, to pass' containing the prefix *ver*-, the stem *geh*, and the infinitive ending -*en*, this results in different values for the individual sounds:

Table 8.A. The calculated information content for the German verb *vergehen* 'to go by, to pass' according to Dellert (2019:108).

[f]	[ɛ]	[ɐ]	[g]	[eː]	[ə]	[n]
0.597	0.228	0.144	0.762	0.615	0.076	0.012

If the initial data are wordforms instead of allomorphs (e.g., *helfen* and *hilft*), this procedure can be used to assign lower probabilities to obviously incorrect correspondences (in the exemplary wordforms *en-t*).

Kondrak and Sherif (2006:48) found that even simpler statistical models and methods in cognate identification yield better results than a simple Levenshtein distance. Nevertheless, such approaches only become interesting for IR when a language has a high degree of suppletion.

8.1.2.3 How to Deal with Suppletion

Suppletion or near-cognates inevitably lead to false correspondences in the procedure. Identifying them requires a procedure comparable to the cognate and loanword detection of the automated Comparative Method. The basic principle is similar to that of historical linguistics, where discord with sound laws is an indication of borrowing or innovation (cf. Sect. 3.2). To make this possible in an automated way requires extremely valid sound correspondences from ideal source data and subtly nuanced models to represent the sound environments accurately enough. A historical linguist also has extra-linguistic information available to identify loanwords, which enables him to "idealize" the source data to a certain extent. Cognate detection is an area of CHL continuing to be explored. An overview of this is provided by Dellert (2019:124–137). For IR, the identification of suppletion and near-cognates only plays a secondary role. Loanwords occur far more frequently than suppletion in word correspondences of two languages. Suppletion may therefore be left behind in the data as "noise," since the focus is primarily on particularly frequent sound alternations. The interest in determining suppletion is relevant to determining extremely old sound changes. Knowing which form of the Latin paradigm *ferro : tuli : latus* is suppletive leads to the identification of an old sound change that cannot be reconstructed internally else. But IR itself offers no possibility to identify suppletion.

8.1.2.4 How to Determine Relevant Correspondences

Relevant are those sound correspondences with unequal sounds. For the paradigm of German *helfen* 'to help', this concerns *e:i* (*helf:hilf*), *e:o* (*helf:holf*), *e:a* (*helf:half*), *i:o* (*hilf:holf*), *i:a* (*hilf:half*), and *o:a* (*holf:half*). For the entire German dataset, this results in a large number of sound alternations, the majority of which do not reflect any sound change. To reduce the number of false alternations, the sound correspondences are evaluated and sorted using a RELEVANT MEASURE. For this purpose, several measures were applied to the untranscribed paradigmatic test data and compared. App. B.21 lists the results for two measures (Jaccard index and a weighted relative frequency) that were best suited to detect historical sound changes. Jaccard index is an established measure for similarity and diversity of sample sets. It takes the ratio of intersection of two sets over the union of both morpheme lists.

$$Jaccard(x,y) = \frac{freq(x,y)}{freq(x) + freq(y) - freq(x,y)}$$

The logarithmic Jaccard index gives an adequate possibility to represent historical sound change. With non-transcribed data, *umlaut* and *ablaut* dominate the best-ranked positions. Phonological rules are mostly not realized orthographically so they only become visible with transcribed source data. With IPA-transcribed data, the feature *stressed:unstressed* dominates in the most relevant positions, which is related to the fact that stress is usually marked solely in polysyllabic words (see App. B.21).

8.1.3 Determination of Conditions

8.1.3.1 Relevance Measure

In order to derive a historical sound law from sound correspondences, the condition of the sound alternation is required. For this purpose, all wordforms of all lexemes with the corresponding alternation are collected and compared (the corresponding sound itself is replaced by an underscore):

- e: { *h_lf-e* (1^{st} sg. pres.), *h_lf-en* (1^{st} pl. pres.), … }
- i: { *h_lf-st* (2^{nd} sg. pres.), *h_lf-t* (3^{rd} sg. pres.), … }

To determine relevance, it is measured which sounds (environment *X*) and sound sequences (environments *_X*, *X_*, and *X_Y* are considered) occur most frequently in word set A relative to word set B. Using PERCENTAGE DISTRIBUTION is one possible measure. App. B.22 (left column) shows the corresponding values for the German allophones [x] and [ç] based on the morphemes having this alternation in their paradigms. The percentage distribution already gives indications for the correct condition ([ç] is used after front vowels), but wrong environments (e.g., ʔ and _p) can also be found, which are caused by the low frequency of these environments. This problem also occurs in other relevant measures, such as the difference measure of relative frequencies, in which the relevant values for sound A represent positive values, and the relevant values for sound B represent negative values:

$$conDiff(c, A, B) := \frac{freq(c)\ in\ wordlist\ A}{\sum_i freq(i)\ in\ wordlist\ A} - \frac{freq(c)\ in\ wordlist\ B}{\sum_i freq(i)\ in\ wordlist\ B}$$

The result shows similar values as the percentage distribution (App. B.22 right column) but allows a simultaneous evaluation of sound pairs, while a percentage measure requires separate considerations of *ç-x* and *x-ç*. A disadvantage of this measure is the inconsistent normalization of the values, which makes it unsuitable to compare the result of different sound pairs. This already becomes clear in the consideration of the value of the most relevant true positives for [x] (-0.0364: ˈɔ, ˈɔ_, and ˈa_). In other sound pairs, this value is clearly in the true-negatives range.

Two conclusions can be drawn from these considerations. First, the sound environment X (i.e., the occurrence of a sound x at any position in the word) is not suitable as a potential condition, since the size of the sets (i.e., the number of words with alternating sounds) is, in practice, too small to be able to assume an approximately equally distributed frequency of non-relevant environments. However, this is necessary to classify them as irrelevant. Second, the random occurrence of an overall low-frequent environment in a set may bias the result. For avoiding this, the use of significance measures may be helpful.

8.1.3.2 Significance Measure

A significance measure is used to determine whether deviating occurrences of an environment in a sound set are statistically significant compared to the total occurrence of this environment. Random deviations can be identified as such and taken into account in the final evaluation. At the same time, however, relevant environments can be sorted out because their occurrence is too small to achieve an expected value μ above the necessary minimum value.

Table 8.B. 2x2-contingency table of a four-field test. The expected value E for each cell is the factor of the row sum and column sum divided by the total number of all four fields.

	Environment Exists	**Enviroment Does not Exist**
Sound A	freq A	freq B
All Sounds	freq C	freq D

In this thesis, the *chi-squared four-fold test* is used, which checks whether the distribution of a feature (in this case, the sound environment) is identical in two groups (in this case, with sound A and the overall distribution). It is only applied to sound environments that yield an expected value above 5, as the test is invalid for smaller expected values (cf. Nelson 2020:168). The result of the test indicates how significant the deviation of the occurrence of the environment is with the sound A compared to the total occurrence (see App. B.23). By comparison with the whole corpus, environments are rated higher if they occur exceptionally frequently in the sound set. However, since the word set of sound A also contains many wordforms or derivations of the same root (e.g., *brechen* 'to break' : *brach* and *unterbrechen* 'to interrupt' : *unterbrach*), many environments that are not relevant for IR are also considered significant. The significance measure can be multiplied by the relevance measure as an indicator function to weed out insignificant values. An indicator function takes the value 1 if a certain condition is met (in this case: if the significance value is above the threshold p), otherwise 0. The function *condMeasure* with the sound pair *A-B* and the condition *c* can thus be defined as follows:

$$condMeasure(c, A, B) := relev(c, A, B) \times \chi_T(c) \ with \ \chi_T(c) = \begin{cases} 1 \ if \ \chi^2 \leq p \\ 0 \ if \ \chi^2 > p \end{cases}$$

To determine which threshold value is appropriate for *p*, the percentage share of the *condMeasure* values of all conditions that are actually (*true*) conditions of the sound pair is calculated:

$$\frac{\sum_i condMeasure(c_i, A, B) \ with \ c_i \ is \ a \ true \ condition \ of \ A - B}{\sum_i condMeasure(c_i, A, B)}$$

High precision can be resulted from a high percentage share. The case $p = 0$ sets the significance measure for all probabilities to 1, so that this value represents the percentage share for the relevance measure without taking significance into account. App. B.26 shows the percentage value with different p-values for three German sound pairs: *ç-x* and *ʀ-ɐ̯* are allophone pairs, *t-d* morphophonemes. All three sound pairs show their maximum at very high p-values (*ç-x*: 0.9; *ʀ-ɐ̯*: 0.95; *t-d*: 0.95), which argues for the use of a significance measure with a high threshold.

Table 8.C. Identified conditions for the German sound pair *ç-x* with a threshold p = 0.95 (second column) and without any threshold (third column). The true positives are indicated in italics.

ConRel(c,A,B)	Without Significance Measure	With Significance Measure
0.973684	*ˈɛ*	→ *ˈɛ*
0.972972	*ˈɛ_*	→ *ˈɛ_*
0.965517	*ˈɛ_n̩*	→ *ˈɛ_n̩*
0.875000	ʀ	→ ʀ
0.833333	*ɪ_, ɪ_#, ˈɛ_t*	→ *ɪ_, ɪ_#, ˈɛ_t*
0.823529	_n̩	→ _n̩
0.800000	*ɛ*	→ *ɛ*
0.777778	ə	
0.769230	l	
0.763157	n̩	→ n̩
0.761904	n	
0.750000	v, *ˈø:, ɪ*, d, ˈʊ	
0.714285	h, _#, ʁ	→ ʁ
0.702127	b	→ b
0.700000	s	

However, as the *p*-threshold increases, the recall also decreases significantly. As a consequence, the number of "identified" conditions decreases with the increasing *p*-value. For a *p*-value of 0.95, there are only ten conditions for the sound pair *ç-x* that show a value above 0.0. That the use of a high significance measure is nevertheless worthwhile is shown in Tab. 8.C. The second column shows the conditions without the use of a significance measure, and the third column with removed conditions by the measure. The majority of the filtered conditions are false conditions.

8.1.4 Transforming to Sound Changes

8.1.4.1 Creating Feature Pairs

Table 8.D. The most relevant sound correspondences (cf. App. B.21) and the proposed conditions with *conRel*-values more than 0.8.

Jaccard(x,y)	Conditions (> 0.8 and p > 0.9)
ˈœː-œː (0.50000)	-
ˈæ-æ (0.22222)	-
ʀ-ɐ̯ (0.13305)	_ə, ə_ə, ə_, _ɐ, d_ə, #_, t_ə, ç_ɐ, k_ə
l̥-l (0.09313)	_t, k_t, d_t, t_t, b_t, g_t, p_t, f_t, m_t, s_t, …
ˈuː-uː (0.07744)	m_t
ˈyː-yː (0.06718)	_ʀ, ə, g, k_ʀ
ʁ-ɐ̯ (0.06608)	_ə, ˈiː_ə, _ˈiː, a_, ˈaː_ə, _a, #_, _e, ˈeː_ə, _i, #_e, …
ˈøː-øː (0.06540)	ə, g, _ʀ
g-k (0.06217)	_ə, _n̩, ˈiː_ə, ˈaː_ə, ˈeː_ə, #_ə, ˈyː_ə, ɪ_ə, ˈøː_ə, …
ˈʊ-ʊ (0.06033)	ə, _ʀ, g

The result of the procedure is a list of relevant sound pairs with suggested phonetic conditions. For the German IPA dataset from App. B.21, the relevant values can be seen in Tab. 8.D. In the final step, both the sound correspondences and the conditions have to be transformed to sound laws by combining concrete sound changes (i.e., sound correspondences with conditions) of the same pattern. For this purpose, the sound correspondences must be replaced by feature correspondences making a list of distinctive features of each sound necessary:

- ˈaː → {unrounded, open, front, vowel , stressed, long}

The choice of distinctive features can strongly influence the result. Apart from different views on individual features (e.g., [a] as a front vowel or neutral), this applies especially to the handling of “negative” features (e.g., *unrounded* as *non-rounded*, *short* as *non-long*). To get a more meaningful result, the number of features is defined as much as possible. Diphthongs and affricates are thus defined with both the features *diphthong* and *vowel* or *affricate* and *consonant,* and with the features of both sound parts.

The FEATURE PAIRS are generated from all sound pairs. The sound pair ˈaː-aː thus produces 6^2 = 36 feature pairs, of which five pairs contain identical features (*unrounded-unrounded, open-open, front-front, long-long*) and are ignored. Similarly, all feature pairs containing both sounds are irrelevant, leaving only the feature pair *stressed-unstressed* for the sound pair *ˈaː-aː*. In this way, 463 feature pairs were created, each of which inherits the wordforms of the parent sound pairs. On the basis of the wordforms, the relevance can be calculated for each feature pair as for any other sound pair. For the determination of the conditions, these must also be decomposed into their distinctive features. For the conditions X (occurrence of the sound in the wordform), X_, and _X (occurrence of the sound immediately before or after the target sound), *n* new environments (from now on "feature environments") result for each previous environment of the sound pair with *n* as the number of distinctive features for X. For the environments of the pattern X, the relevance is calculated as for any other sound correspondence. For the environments of the pattern X_Y, $n \cdot m$ feature environments with *m* as the number of distinctive features of Y apply. For the German dataset alone, this has resulted in over 60,000 feature environments.

In order to be able to eliminate duplications of conditions, subsets are identified and stored as such at this point. For the hypothetical feature environments *dental_dental* = {t_t, d_t, t_d} and *plosive_plosive* = {t_t, d_t, t_d, p_t}, *dental_dental* ⊆ *plosive_plosive* holds. The new feature environments are finally evaluated with relevance and significance measures just like the environments of sound pairs.

Table 8.E. Example of newly formed feature environments generated through the German wordform [klaɪ̯nʃt_t] (for [klaɪ̯nʃtat] 'small town' and [klaɪ̯nʃtaːt] 'small state').

k	→ velar, plosive, voiceless
l	→ lateral, alveolar, …
a	→ …
…	
t_	→ dental_, plosive_, voiceless_
_t	→_dental, _plosive, _voiceless
t_t	→ dental_dental, dental_plosive, dental_voiceless, plosive_dental, …

8.1.4.2 Reducing Feature Pairs

In the next step, the most relevant sound changes have to be determined. In addition to "abstract" sound changes (e.g., C[+voiced] > C[-voiced] / _#), concrete sound changes such as *x* > *ç* may occur in language history. Since these must be described completely or partially without phonetic features, it is advisable to leave both feature pairs and sound pairs in the process. The result of the procedure is a much longer list containing a variety of doublings (see App. B.24). The preliminary result of App. B.24 shows a large number of values among the ten most relevant correspondences which reflect the sound correspondence ʀ-ɐ̯. To avoid these duplications, an iterative procedure is applied, where the most relevant correspondence is removed after each step and the procedure is repeated completely. A feature pair (e.g., *trill-central* at rank 2) is preferred over a sound pair (e.g., *ʀ-ɐ̯* at rank 1), if the relevance values of the sound pairs of the feature pair (*ʀ-ɐ̯* and *ʀ-ɐ* for *trill-central*) do not differ significantly among themselves. Thus, in the correspondence from App. B.24, the sound pairs *ʀ-ɐ̯* and *ʀ-ɐ* are omitted in the second iteration. This means that the recurrent correspondences such as *uvular-near_open* or *uvular-semivowel* are still included in the process, but due to the missing sound pairs *ʀ-ɐ̯* and *ʀ-ɐ*, they are evaluated significantly worse. In this way, doublings can be removed without removing other potentially relevant sound pairs and feature pairs.

After each round, one correspondence with its environments is determined to be relevant. To reduce the environments, a similar procedure is used: The most relevant condition (e.g., *vowel_vowel*) is determined and all further environments that represent a part of it (e.g., *vowel_*) or contain the same environments or a subset of them (e.g., *front-vowel_vowel*) are removed. The results after ten iterations can be seen in Tab. 8.E (for a complete list, see App. B.11).

Table 8.E. Results of the paradigmatic method after ten iterations. A more specified list can be found in App. B.11.

	Feature pair	Jaccard	Sound pair(s)	Conditions (> 0.9)
1	vibrant-central / near_open	0.755414	ʀ-ɐ̯, ʀ-ɐ	ˈiː_ə, V|_|V, front|_|V, unrounded|_|V, …
2	uvular-near_open	0.678119	ʁ-ɐ̯, ʁ-ɐ	ˈiː_ə, V|_|V, front|_|V, unrounded|_|V, …
3	mid_vowel-near_open	0.653292	ə-ɐ	_ʀ, _|uvular, voiced|_|voiced, plosive|_|vibrant, …
4	short-lateral / mid_vowel-lateral / approximant	0.651980	ə-l̩	l, _|fricative, bilabial|_|fricative, …

<table>
<tr><td>5</td><td>voiceless-voiced</td><td>0.610616</td><td>t͡ʃ-d͡ʒ, m-f, ʔ-b, ʔ-l, t͡s-d, z-s, g-f, g-k, ʀ-t, ʀ-s, …</td><td>ˈeː_t, ˈaɪ̯_t, n_#, ˈaɪ̯_#, ˈiː_#, ˈaɪ̯_ə, ˈaʊ̯_t, …</td></tr>
<tr><td>6</td><td colspan="3">ʁ-ʀ</td><td>ˈa_t, _#, ˈɪ_t, open|_|plosive, open|_|voiceless, …</td></tr>
<tr><td>7</td><td>stressed-unstressed</td><td>0.590648</td><td>ˈɛ-ɪ, ˈa-eː, ˈɔ-eː, ˈaɪ̯-eː, ˈoː-ɔɪ̯, ˈiː-iː, ˈœ-œ, ˈɛ-ɛ, …</td><td>_b, _ʀ, plosive|_|unstressed, plosive|_|short, …</td></tr>
<tr><td>8</td><td>close-unrounded</td><td>0.599788</td><td>ˈuː-ˈaː, ˈiː-ˈaɪ̯, ˈyː-ˈaː, ˈɛː-ˈuː, iː-aɪ̯, ˈyː-ˈa</td><td>n̩, #_b, #_, ʁ_b, t_g</td></tr>
<tr><td>9</td><td>front-back_vowel</td><td>0.597482</td><td>ˈøː-ˈuː, ˈɛː-ˈoː, ˈyː-ˈuː, ˈɔ-ˈøː, ˈʊ-ˈi, ˈʏ-ˈuː, …</td><td>_ç, t͡s_g, _χ, χ, _k, _ɐ̯, ç, n̩, t, velar|_|semi-vowel, …</td></tr>
<tr><td>10</td><td>close_mid-close</td><td>0.561221</td><td>ˈøː-ˈyː, ˈøː-ˈiː, ˈuː-ˈoː, iː-eː, ˈiː-ˈeː</td><td>_#, z_#, t͡s_g, _k, k, l_k, voiceless|_|velar</td></tr>
</table>

8.2 Evaluation

For the morphophonemic methods, all four evaluations presented in Sect. 7.5 are applied. In order to assess the effect of transformation, the match rate and directional evaluation are performed for both transformed and single-sound alternations. The evaluation of the complementary pairs and conditions is largely restricted to the transformed sound changes. A temporal evaluation taking into account the language stages of German should give an answer as to how old the internally reconstructed sound changes are.

8.2.1 Match Rate

The automatically identified sound pairs are assigned to one of the following categories for evaluation:

- PHONOLOGICAL RULE (PR) and SOUND CHANGE (SC)
- FREE VARIATION (FV): The pair includes free variants (e.g., *ʀ-ʁ*). This includes phonological rules and sound changes which otherwise occur twice in the dataset due to free variants (*ʀ-ɐ̯* in addition to *ʁ-ɐ̯*).
- MORPHOLOGICAL ALTERNATION AND SUPPLETION (MA): These include alternations that exist synchronically but are not based on a historical sound change or phonological rule (e.g., *ablaut*).
- TRANSCRIPTION ERRORS (TS): These alternations are due to errors in the source data or transcription-related alternations (e.g., the marking of stressed vowels in polysyllabic words).
- NOISE: This includes sound pairs that are based on incorrect assignments of the algorithm or implementation.

If a sound pair can be assigned to more than one category, it counts towards the "higher" category in the evaluation. The categories *PR/SC* and *FV* are considered "true positives." Since free variation does not occur in the gold standards but are more or less due to sound change, they are ignored in the evaluation and considered in a separate evaluation (PRECISION WITH FV). The same applies to "doublets", which are the result of a phonemic split. In the "machine evaluation", all categories except "noise" are considered "true positives" (MACHINE PRECISION). This is to determine how many errors can be attributed to the algorithm or the data set, respectively.

8.2.1.1 Single-Sound Alternations (Paradigmatic Approach)

Table 8.F. Results of the paradigmatic method after fifteen iterations (single-sound). Legend: Gold standard 1 (see App. A.1), Gold standard 2 (see App. A.2). P/S: phonological rule and sound change, PR: phonological rule, SC: sound change, FV: free variation, Dubl: doubletted phonological rules and sound changes, Abl: ablaut. R stands for any sonorant; I for *i, ī, j*, or *iu*.

SP	Example		Gold Standard 1	Gold Standard 2
1 ʀ-ɐ̯	[ɛɐ̯ˈhøːʀə : ɛɐ̯ˈhøːɐ̯t]	P/S	ʀ > ɐ̯ / _C_0	r > ɐ / _$
2 l̩-l	[plaʊ̯ˈziːbl̩ : plaʊ̯ˈziːblɐ]	P/S	l > l̥ / .C_0_C_0.	(N) ʁ > ɐ / _$
3 ʁ-ɐ̯	[ˈmeːʁəʁə : meːɐ̯]	FV	see 1 and 8	see 1 and 8

4 g-k	[ˈtaːɡə : taːk]	P/S	g > k / _[-R]$_0$.	g > p / _#
5 p-b	[loːpt : ˈloːbn̩]	P/S	b > p / _[-R]$_0$.	b > p / _#
6 ə-ɐ	[ˈnaːə : ˈnɛːɐ]	PR	ə > 0 / _C[+R]C$_0$.	-
7 ə-l̩	[ˈklɪŋlə : ˈklɪŋl̩n]	Dubl		
8 ʁ-ʀ	[ˈiːʁə : ˈiːʀɐ]	FV	-	-
9 ʀ-ɐ	[ˈɡaɪ̯stəʀə] : [ˈɡaɪ̯stɐt]	PR	ʀ > ɐ / .C$_0$_C$_0$.	-
10 ˈʊ-ʊ	[ˈʁʊndə : ʁʊnt]	TS	-	-
11 ˈɛː-ˈeː	[ɡeːbn̩ : ˈɡɛːbn̩]	Abl	-	-
12 ˈɪ-ˈɛ	[bəˈʃpʁɪçt : bəˈʃpʁɛçt]	SC	-	*e > *i / _NC, C$_0$I
13 z-s	[ˈleːzn̩ : leːst]	P/S	z > s / _[-R]$_0$.	z̥ > z / X_V, #_V z > s / _#
14 ˈøː-ˈoː	[ˈʁøːtɐ : ˈʁoːtə]	SC	-	ˈō > ˈœ / _...I
15 ˈɔ-ˈœ	[ˈtɔxtɐ : ˈtœçtɐ]	SC	-	ˈo > ˈö / _...I

Tab. 8.F shows the result of the first 15 sound pairs with the corresponding categorization. A list of the 100 best-rated sound pairs can be found in the App. B.1. The majority of the identified sound pairs in the upper range (> 0.64) can be assigned to a historical sound change or a phonological rule. Fig 8.3 shows the percentage of categories between 0.3 and 0.8 in 0.05 steps. A high proportion of true positives (*PR/SC* and *FV*) can be seen in the upper range, but this sharply drops off at 0.64. In the middle range (0.45–0.65) transcription errors and morphological alternations dominate. True errors (*Noise*) only occur in the lower range. The solid black line in Fig. 8.3 indicates the *precision value with FV*. This makes clear that morphophonemic IR as a method only maps historical sound change in the upper range. False alternations frequently occur at values lower than 0.6. However, a large part of these can be attributed to automation (i.e., incorrect data or incorrect sound correspondences). The dashed line in Fig. 8.3 marks the machine precision.

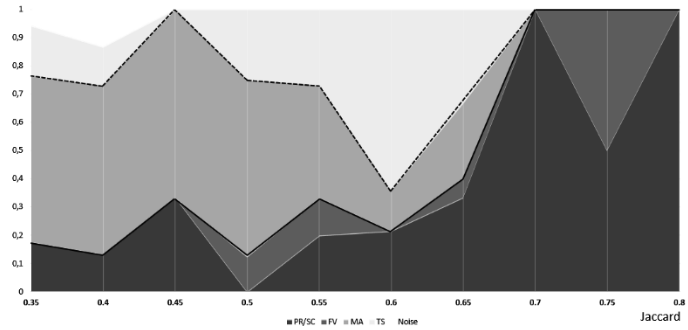

Figure 8.3: Evaluation of the paradigmatic approach (single-sounds): Percentage of the four evaluation categories for different Jaccard indices between 0.35 and 0.8. The individual values were summarized with the next higher 0.05 Jaccard value.

Due to the free variations of *ʁ-ʀ* (rank 8) and *χ-x* (39), some sound changes (3, 45, 53) occur repeatedly. Nevertheless, in principle, free variation is recognized and does not cause fatal errors. In total, there are five cases of "doublets" among the first 100 sound pairs (i.e., they represent the same sound change). Three of them (41, 51, 98) are due to vowel lengthening in open syllables preceded by another sound change: The alternation *ˈɪ-ˈeː* (e.g., [tʁɪt͡st] 'you (sg.) step' : [ˈtʁeːtə] 'I step') arose by the raising of short ɛ when there was an *i*-sound in the following syllable — this pair was identified in 12 (*ˈɪ-ˈɛ*). The subsequent lengthening of ɛ in open tone syllables has created the second sound pair *ˈɪ-ˈeː*. The lengthening (*ˈɛ-ˈeː*) itself was not identified. A mismatch is caused by sound changes with zero, such as *ə* > 0 / _C[+sonorant]C0. (6 and 7). These appear as lost sound_subsequent sound in Tab. 8.F (*ə-ɐ* and *ə-l̩*), as the Levenshtein alignment did not identify the syllabic and non-syllabic sonorants as belonging together (e.g., [ˈʁeːgl̩n] ‚to manage' : [ˈʁeːgələ] ‚I manage'). More complex sound-alignment algorithms may handle loss-sound correspondences in a more appropriate manner. The majority of transcription errors stem from the practice of not setting stress on monosyllabic words. This results in sound pairs of the type *ˈʊ-ʊ* (10), which account for a large proportion of the sound pairs in the middle range. Their occurrence does not have a negative influence on the evaluation, but they may cover genuine historical stress change. These errors can be corrected by adjusting the source data. At the back of the list, there are also transcription errors caused by the editors of Wiktionary. For instance, the alternation *p͡f-f* was caused by the different transcription of *p͡f* as *pf*.

In the middle range, morphological alternations occur more frequently, which are mainly due to ablaut. The Proto-Indo-European ablaut *e:o:Ø* has multiplied through secondary sound changes in German (e.g., *o > a > ɛ, e > i*): [ˈneːmən] ‘to take’ : [nɪmt] ‘takes (3rd sg.)’ : [naːm] ‘took (1st and 3rd sg.)’ : [ˈnɛːmə] ‘would take (3rd sg.)’ : [gəˈnɔmən] ‘taken’. Thus, besides the historically correct ablaut correspondences (in the example: *eː - aː - ɔ*), sound pairs arose which can be attributed neither to historical sound change nor to ablaut (*eː-ɛː, aː-ɪ,* or *ɔ-ɪ*). These sound pairs inevitably arise with the morphophonemic method and are excluded in “manual” IR by linguists by means of inferring methods. Purely internally, however, it is often difficult to decide which of the sound alternations in the paradigm is historical. For example, the presence of the alteration *e-ɪ* in the present tense paradigm argues for it being a historical sound alternation. At the same time, however, a similar argument would lead to the historical incorrect pairing of [ˈneːmən]:[ˈnɛːmə]. The sound pair *ˈeː-ŋ*, which can be traced back to the suppletion paradigm of [ˈgeːən] ‘to go’ : [gɪŋ] ‘went’ : [gəˈgaŋən] ‘gone’ is also considered a “morphological alternation” in this evaluation. There are many derivations of this verb, which make the sound pair *ˈeː-ŋ* frequently occur in the data.

Noise occurs especially in the lower range (< 0.37), where the procedure could not determine the correct sound correspondence (e.g., ˈiː-g through [ˈt͡siːən] ‘to pull’ : [gəˈt͡soːgn̩] ‘pulled (participle)’, or ˈɛː-t through incorrect morpheme analysis of word roots ending with *-t/-d*, such as [ˈtʁeː-tn̩] ‘step (3rd pl.)’ : [tʁaːt] ‘stepped (3rd sg.)’).

8.2.1.2 Single-Sound Alternations (Derivational Approach)

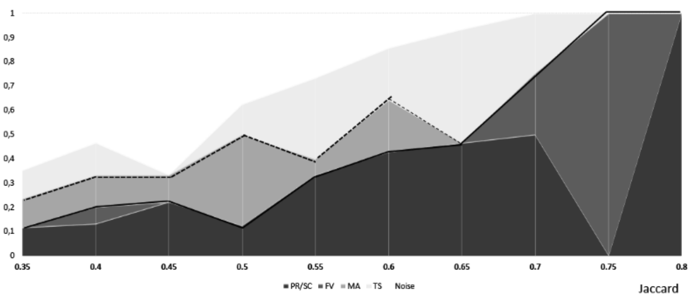

Figure 8.4: Derivational approach: Percentage of the four evaluation categories for different Jaccard indices between 0.35 and 0.8. The individual values were summarized with the next higher 0.05 Jaccard-value (see Fig. 8.3 for the paradigmatic approach).

The derivational approach has achieved slightly better result than the paradigmatic approach. App. C.1 shows the single-sound evaluation of the derivational method.

It shows generally lower Jaccard indices, which can be explained by the lower recurrence of an alternation in the derivations compared to the paradigmatic forms. As can be seen from the comparison between the percentage values of Fig. 8.3 and Fig. 8.4, the proportion of correctly identified sound alternations is higher in the middle range (0.55–0.7) than the corresponding value of the paradigmatic approach. The percentage of morphological alternations decreases due to the lower importance of the *ablaut* in derivational formation. The sound correspondence *t͡s-t* or *t͡s-d* comes from word pairs such as [apˈstʁakt] 'abstract' : [apstʁakˈt͡si̯oːn] 'abstraction' and [mɪˈli̯aʁdə] 'billion' : [mɪˈli̯aʁt͡stl̩] 'part in a billion'. It describes a phonological rule *t* > *t͡s* / *_s*>0, which, however, is not listed in the gold standard. It is therefore evaluated as a simple morphological alternation. At the same time, the proportion of misidentifications already occurs at high Jaccard indices and increases sharply as the Jaccard index decreases, so that the majority of identifications with Jaccard indices lower than 0.45 are errors. This trend is unsurprising since derivational morpheme analysis increasingly leads to false correspondences. For instance, the word pairs [demokʁaˈtiː] 'democracy' : [demokʁatɪ-ˈz-iːʁən] 'democratize' and [demoˈkʁaːtɪʃ] 'democratic' : [demokʁati-ˈz-iːʁən] lead to the false correspondence *ˈiː-ɪ* and *ʃ-z*, respectively.

In total, 28 phonological rules or sound changes were detected among the first 100 pairs — three more than with the paradigmatic approach. The newly identified rules are shortenings of long vowels in unstressed syllables (e.g., [ˈʁeːgl̩] 'rule' : [ʁeguˈlɛːɐ̯] 'regular'). Since Germanic-inherited words usually have fixed stress on the word stem and the rule is almost exclusively applied in loanwords; it does not appear in paradigms. For the same reason, the rule is absent for *umlauts*, which occur less frequently in loanwords. Similarly, the sound change *x* > *k* / _s has been detected by the word pair [zɛks] 'six' : [ˈzɛçt͡sɪç] 'sixty'; in the paradigms, this sound change has been levelled (e.g., [ˈlaxn̩] 'to laugh' : [laxst] 'laugh (2nd sg.)'). The derivational approach has not been able to identify the pairs of syllabic sonorants (*n* > *n̥* and *l* > *l̥*), which have little importance in derivational formation.

8.2.1.3 Temporal Evaluation

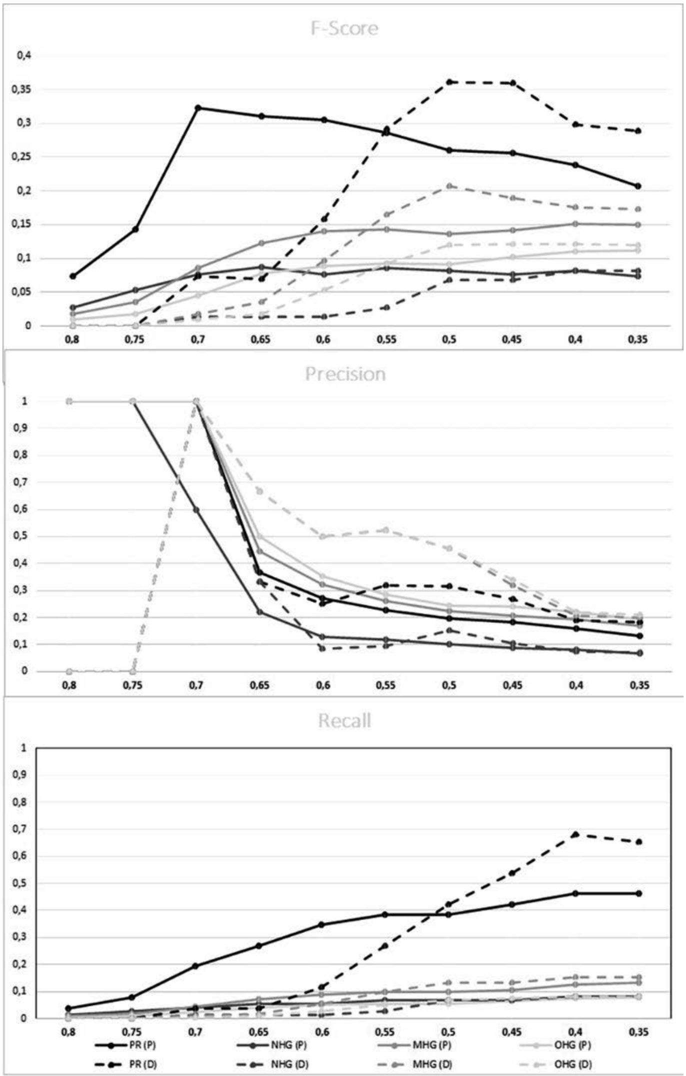

Figure 8.5: Temporal evaluation with different Jaccard indices used as thresholds. The solid lines indicate the paradigmatic approach, the dashed lines the derivational approach. The phonological rules (black, PR) achieve the best F-score and precision values. At the lower Jaccard values, Middle High German (grey, MHG) overtakes the values of New High German (dark grey, NHG).

For the temporal evaluation, the free variants and duplicates are ignored since they complicate the calculation of recall values. Fig. 8.5 shows the F-score, precision, and recall values for phonological rules, (Pre-)Old High German (OHG), (Pre-)Middle High German (MHG), and (Pre-)New High German (NHG) sound changes (for details, see App. B.3–B.5 and C.3–C.5). The values of the older language stage include the sound changes of the younger stages. The Jaccard scores are divided into 0.5-intervals.

As already described in the previous section, the precision value decreases proportionally to the threshold, while the recall increases as expected. The precision increases with each additional time step since more sound changes are added. The reason why the phonological rules have higher precision and recall values than the NHG sound changes is that the German final devoicing, although a productive rule in German, dates from Pre-Middle High German times. This leads to the conspicuousness that the MHG-scores overtakes those of NHG at low threshold values, contrary to expectation. The recall, which indicates how many of the sound changes could be identified in the gold standard, reaches a maximum of 0.1518 for the sound changes, while half of all phonological single-sound rules were recognized. These observations support the view that the morphophonemic approach primarily reconstructs phonological rules.

The derivational approach shows the highest F-scores at thresholds between 0.45 and 0.5, which is due to a variation of stress position among the derivational forms. In comparison to the paradigmatic approach, the derivational approach shows higher optimum scores for phonological rules (0.3607 vs. 0.3226), as well as for MHG (0.2069 vs. 0.1514). The fact that this approach performs better than the paradigmatic approach for phonological rules is related to the German peculiarity that the shortening of unstressed vowels is nearly only observed in derivations. For sound changes, the observation can be made that the paradigmatic approach performs better for the younger language stages and the derivational approach for older stages. This can be explained by the pressure of paradigmatic levelling, as can be seen in the example of *x* > *k* /_s already mentioned. The oldest identified sound changes are from the Germanic period: the Early Germanic raising of *e* and the nasal loss before *x*.

8.2.1.4 Transformation

Tab. 8.G shows the result of the paradigmatic approach with transformed correspondence pairs. A complete list can be found in App. B.11. Fewer pairs are found through merging of the single-sound changes and the procedure is completed after 34 iterations. As already discussed in the literature, the first identified sound changes represent synchronic sound alternations. Among the first seven iterations, five phonological rules from the gold standard are found, with the highest precision value being at 0.8 (with a threshold of 0.6). The two incorrect sound changes are the free variation of *ʀ-ʁ-r*. The words with /r/ were transcribed differently in the source data, which also resulted in a duplication of /r/ > ɐ / _C_0. The gold standard offered a total of nine automatic alternations for German. Of these, six were identified, with the highest recall being at 0.667 (with a threshold of ≤ 0.45).

Table 8.G. Results of the paradigmatic approach with transformed feature pairs after fifteen iterations (for the complete result, see App. B.11; for the meaning of the abbreviations, see Tab. 8.F).

Feature Pair	Sound Pairs		Gold Standard 1	Gold Standard 2
1 vibrant-central	ʀ-ɐ̯, ʀ-ɐ	PR	ʀ > ɐ̯ / _C_0	r > ɐ / _$
2 uvular-near_open	ʁ-ɐ̯, ʁ-ɐ	PR	*see (1)*	(N) ʁ > ɐ / _$
3 mid-near_open	ə-ɐ	PR	ə > 0 / _C[+R]C_0.[1]	ə > 0 / _ɐ *or* r̥ > ɐ
4 mid-approximant	ə-l̩	PR	R > [+syllabic] / C_0_C_0[2]	R > R̥ / ə>0_
5 voiceless-voiced	ç-g, p-b, t-d, t-b, …	PR	[-R] > [-voice] / _[-R]	D > T / _#
6 ʁ-ʀ	-	FV	*free variants*	
7 stressed-unstressed	ˈɛ-ɪ, ˈa-eː, ˈɔ-eː, …	TS	-	-
8 close-unrounded	ˈuː-ˈaː, ˈiː-ˈaɪ̯, …	Abl	-	-
9 front-back	ˈøː-ˈuː, ˈɛː-ˈoː, …	SC	-	*umlaut*
10 close_mid-close	ˈøː-ˈyː, ˈuː-ˈoː, …	SC	-	-
11 near_close-open_mid	ɪ-œ, ˈɪ-ˈɛ, ˈɔ-aʊ̯, …	SC/ Abl	-	*e > *i / _NC, C_0I
12 short-unrounded	ˈʏ-ˈeː, ˈɪ-ˈaɪ̯, ɔɪ̯-ɐ, …	Abl	-	-
13 χ-ç	-	PR	/ç/ > [-front] / V[-front]_	x > ç / {-V[front]}_

14 short-long	ˈɛ-ˈaː, aː-a, ˈoː-ˈo	SC/ Abl/TS	V > [-long] / _[-stress]	ˈVː > ˈV /__ht,rC
15 ˈɛː-ˈeː	-	Abl	-	-

The spirantization /g/ > [ç] / /ɪ/[-stressed]_C_0 was also recognized but attributed to the final devoicing (see sound pairs of It. 5 in Tab. 8.G). A phonological rule not detected by the implementation was the glottalization 0 > [ʔ] / (_V...)ω. This rule is barely detected morphologically. The second unrecognized alternation was the nasal assimilation, as in *haben* [haːbm̥] 'to have' and *legen* [leːgŋ̊] 'to lay'. These took place in grammatical morphemes, whereas this procedure dealt exclusively with word stems.

The temporal evaluation with transformed sound changes is listed in App. B.15–B.15 and C.13–C.15. Despite the incorrect mappings made by the transformations, the best F-score for phonological rules increases significantly compared to the single-sound evaluation (PR, paradigmatic: 0.5 > 0.3226). In contrast, the evaluation of the historical sound changes shows a tendency towards a slight deterioration (NHG: 0.16 < 0.2, MHG: 0.1968 < 0.2128). The increase in Old High German (0.1516 > 0.1244) may be explained by the newly identified sound change *z > r* (e.g., *waren* 'were' : *gewesen* 'been'), which was under-rated in the single-sound evaluation. The different development of phonological rules and sound changes is another indication that the morphophonemic approach tends to identify synchronic alternations.

The derivational approach yields more sound changes. As already observed for the single sounds, the result of the derivational approach is significantly improved by the detection of older sound changes. The phonological rule of assimilation of syllabic nasals was additionally identified, which, however, is rather caused by an inconsistent transcription in the source data.

8.2.2 Evaluation of the Sound Correspondences

Table 8.H. Evaluation of the sound correspondences generated through the paradigmatic approach with transformed sound changes. The precision values indicate how many of the single-sound changes belong to the sound change of the gold standard.

Feature Pair	Gold Standard	Precision	Recall	F-Score
vibrant-central *or* near_open	ʁ > ɐ / _$	0.5	1	0.666667
mid-near_open	R > R̥ / ə→0_	0	0	-

voiceless-voiced	D > T / _#	0.25	0.857143	0.387097
front-back	*umlaut*	0.147059	0.555556	0.232558
near_close *or* centralized-open_mid	*e > *i / _NC, C_0i,j,u	0.1	1	0.181818
χ-ç	x > ç / {-V[-front]}	0	0	-
short-long	ˈV: > ˈV /__ht,rC	0.147059	0.555556	0.232558
long-velar	ŋ > 0 / V>V:_h,x	0	0	-
ʀ-z	*z > *ɹ̯ / _	1	1	1
TOTAL		0.166667	0.542857	0.255034

The analysis of the first 100 sound pairs shows that of twelve correctly identified phonological rules, the paradigmatic approach has correctly identified eleven sound correspondences (0.9167, see App. B.6). In contrast, 14 of 17 identified sound changes (0.8235) are correct complementary sounds. With the derivational approach, these values were lower (PR: 0.7647 and SC: 0.5556, see App. C.6). A source of this error is secondary sound changes that can mask the correct sound pair. A good example is the sound change *ou* > *öu* [ˈœʏˀ] / _*...i,ī,j,iu*, which has produced the synchronic sound alternation *ɔɪ̯-aʊ̯* by the secondary sound changes *ou* > *au* and *öu* > [ɔɪ̯]. Since the probability of secondary sound changes increases with time depth, the derivational approach is most affected here. The second source of error is the incorrect determination of complementary sounds and conditions, which is especially observed for a loss (e.g., the sound pair *ə-n̥* reflects the sound change *n* > *n̥* / *ə>0_*).

The precision and recall scores are comparatively poor for the evaluation of complementary sounds. An F-score of 1.0 is only achieved for non-transformed sound changes. This is mainly due to the difficulty of correctly determining the crucial features of sound changes. The feature pair *voiced-voiceless* is too general and also includes *ʀ-s*. The handling of the null morpheme also forms a problem. The Levenshtein matrix identified the sound correspondences *ə:ɐ / r:0* and *ə:l̥ / l:0* from the pairs *CərV : Cɐ#* and *CəlV : Cl̥#*, respectively. A correct assignment would have been possible with additional phonetic knowledge. However, the morphophonemic IR alone does not provide an option here. The results for the derivational approach are much worse (see App. C.16). This is mainly explained by the fact that the method excludes sound pairs if they have already occurred in the best feature pair in a previous iteration.

8.2.3 Evaluation of the Direction of the Sound Change

Different approaches were tested to determine the sound change direction (see App. B.7–B.8, B.17–B.18, C.7–C.8, and C.17–C.18). In the case of synchronic rules, the direction was correctly determined in 41.667% of the cases when the "restricted sound" is considered as a pre-phoneme (PHONETIC DISTRIBUTION, see Sect. 4.1.1.4). The correctly determined directions include the sound correspondences of final devoicing, which also serves as an example for this approach in the literature. According to Chafe (1959:247), the younger sound is the one that is ARTICULATORILY CLOSER TO THE SOUND ENVIRONMENT. In order to test this assumption, only the best-rated environment was selected as the condition. This method identified the correct direction five times out of twelve synchronic alternations. A problem arises here when no clear decision can be made (e.g., is [ə] or [ɐ] closer to [ʁ]?) or the conditions cannot be correctly determined (e.g., for *umlaut*). However, in some cases, the basic assumption is also wrong. For example, the best-rated condition of *ʀ*-*ɐ* is *V_V* and thus closer to *ɐ*, but this condition is an environment of *ʀ*. The third possibility to determine the pre-sound (PHONETIC PLAUSIBILITY) is difficult to capture objectively. One often cannot clearly distinguish between "condition" and "counter-condition" (i.e., the condition of the other sound change direction). Final devoicing *g* > *k* leads to the counter-condition *_V*, which is not formally different from an intervocalic lenition *k* > *g*. If the counter-condition was identified as a condition, this also leads to errors with the restricted-sound approach. This circumstance leads to a strong fluctuation of the results, ranging from 0.1667 to 0.75 (see App. B.7 and C.8), making them of little use for practical purposes. In my estimation, only two of all identified directions (*ʀ* > *ɐ* or *ʀ* > *ɐ̯*) can be considered more plausible.

The comparison of these approaches shows that the restricted-sound approach and the articulatory-proximity approach show very similar values but could predict only every second direction correctly. The phonetic-plausibility approach, on the other hand, shows high precision but is not applicable to most cases. The direction of a phonological rule and its corresponding sound change may differ. Historically, for example, the voiced [z] in German arose at a word-initial position and between vowels from *s*. The modern phonological rule, on the other hand, requires a change direction *z* > *s*. Another example is the pair *x* and *ç*: synchronically, [ç] tends to occur as the "unmarked" allophone (e.g., in O'Brien and Fagan 2016:50).[61] Surprisingly, the historical sound changes show slightly better results.

8.2.4 Evaluation of the Conditions

Table 8.I. Evaluation of the conditions of five correctly identified transformed pairs (phonological rules, paradigmatic approach). To hide the influence of an incorrect change direction, the conditions (Con.) of the gold standard were adjusted when the automation has determined an incorrect direction (adjusted condition, AC). The percentage of sub-conditions among the true positives is given in the column “doublets.”

Feature Pair	Gold Standard	Con.	Precision	Doublets
vibrant-central *or* near_open	/ʀ/ > ɐ / _C_0 (AC: _V)	66	0.863636	0.982456
mid-near_open	ə > 0 / _C[+sonorant]C_0 (AC: _/r,ʁ,ʀ/)	21	0.714286	0.866667
voiceless-voiced	[-sonorant] > [-voice] / _[-sonorant]$_0$.	109	0.385321	0.928571
χ-ç	/ç/ > [-front] / V[-front]__ (AC : V[-front]_)	26	0.346153	0.888888
short-long	V > [-long] /_[-stressed]	17	0	-
TOTAL		239	0.514644	0.916645

Tab. 8.I shows the result of the evaluation for the phonological rules of the paradigmatic approach with a precision of 0.5146 (see App. B.9–B.10 and B.19–B.20). The precision value indicates the proportion of identified conditions that are a sub-condition of, or identical to, the actual conditions. The doubling rate indicates how many of the identified conditions are merely sub-conditions. In all identified conditions, the proportion of recurrent conditions is very high with all approaches. Since the

[61] In contrast to this, a different view is taken by other linguists (e.g., van Lessen Kloeke 1982:56) who regard [x] as unmarked.

sound change direction was already evaluated in the previous section, the "counter-condition" was used instead of the gold standard's condition for the evaluation when an incorrect direction has been determined (see column "Adjusted condition" in Tab. 8.I; for details on the procedure see Sect. 7.5.1.4). The considered condition patterns (*X_, _X, X_Y,X*) represent adequate means for most rules. Only for one pair (*long-short*) did the case occur that the phonological rule (shortening of unstressed long vowels) could not map the correct condition and, therefore, only wrong conditions were reconstructed. Lower are the precision values for the derivational approach (see App. C.9–C.10 and C.19–C.20). In tendency, the paradigmatic approach seems to be better for reconstructing conditions, as the derivational approach shows a higher error rate due to the large number of composites. The better results of the single-sound evaluation result from erroneous fusions of feature pairs that are formulated too generally. For instance, the sound pairs of final devoicing show very good precision values in the single-sound evaluation. By the transformation to *voiced-voiceless*, wrong conditions from pairs such as *g-t* get into the evaluation and push the value down in this way.

Table 8.J. Comparison of the precision values for phonological rules, sound changes and the diachronic sound changes that are no longer productive as phonological rules. Most of the identified sound changes are still present in the language as phonological rules. "cond." stands for the number of conditions.

	Phonological Rules		**Sound changes**		**Diachronic SC**	
	cond.	precision	cond.	precision	cond.	precision
paradigmatic (single-sound)	216	0.856481	185	0.789189	32	0.3125
paradigmatic (transformed)	239	0.514644	289	0.429066	69	0.014493
derivational (single-sound)	78	0.487179	63	0.571429	4	0.000000
derivational (transformed)	361	0.207756	429	0.174825	164	0.006098

The evaluation shows for sound changes, likewise, lower precision values, which are caused above all by sound changes which do not represent, at the same time, a phonological rule. Tab. 8.J compares the results for the phonological rules and sound changes. Especially for very old alternations the historical conditions are no longer preserved and thus cannot be reconstructed synchronically (column "Diachronic Sound Changes"). This circumstance also suggests that the paradigmatic and derivational approaches are primarily used to reconstruct phonological rules.

9. Semantic Method

The semantic approach is a special case of the morphophonemic method, in which the potential cognates are only related by a loose semantic similarity. This method aims to capture older sound changes from derivations and compounds that have become unproductive, such as *fox-vixen* or *to give-gift*. The disadvantage of this approach is the expected high error rate due to coincidental similarity (e.g., *isle-island*).

9.1 Resources

Semantic similarities can be integrated and processed automatically using semantic ontologies. Ontologies are semantic networks in which concepts are connected by their semantic relation (Fig. 9.1). In common ontologies, such as WordNet (Miller 1995), GermaNet (Hamp and Feldweg 1997, Henrich and Hinrichs 2010) or VerbNet (Schuler 2005), mostly synonymy and hierarchical relations (hypernyms and hyponyms) are represented; other relations are underrepresented. Ontologies are usually designed for a specific language. A multilingual semantic network of adequate scope is provided by BabelNet (Navigli and Ponzetto 2010).

The freely available ontology OntoWiktionary (Meyer and Gurevych 2012) was used for this work. It is based on the data available from the German dictionary Wiktionary. For each concept, hypernyms and other relations are extracted and a *bag of words* is formed containing semantically related words for concepts such as CAT → TOMCAT, TIGER, PET, LION etc.

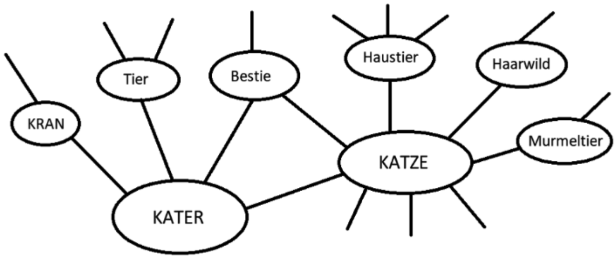

Figure 9.1: Illustration of the concepts Kater 'tomcat' and Katze 'cat' from the German ontology OntoWiktionary. In this depiction, the concepts are ordered by their hypernyms.

The list of these “potential cognates” may become extremely extensive depending on the concept, the relations considered and the hierarchical depth, which also increases the probability of misidentified cognates. The choice of hierarchical depth and relations can thus directly influence the result.

9.2 Detection of Internal Cognates

In order to detect internal cognates, a bag of words consisting of hyponyms, co-hyponyms and their hypernyms is generated for each concept in question. For the concept CAT, this bag of words contains not only numerous predatory and domestic cat species (TIGER, LION, MANX) but also other concepts such as TOMCAT, KITTEN or SPHINX. For the majority of the words, an etymological relationship with the word CAT cannot be assumed, so that a variant of partial cognate detection must be applied at this point. The term “partial cognacy” is used when only parts of a word pair are related, usually the word stem or component stem. List et al. (2016) used algorithms for network partitioning, developed for cognate detection, for the detection of partial cognates. Their study found that “the algorithms generally loose accuracy in the task of ‘true’ partial cognate detection,” (2016:60) but there is no striking tendency.

To identify internal cognates within the bag of words, the phonetic similarity of the words is calculated using a weighted sequence alignment (Jäger 2015:12755–12757). By this method, all word pairs in the bag are aligned using an iterative Needleman-Wunsch algorithm. The Needleman-Wunsch algorithm (Needleman and Wunsch 1970, specified by Waterman et al. 1976) is a procedure that computes the optimal similarity score for each sound correspondence in a matrix by extending the Levenshtein distance method. It uses dynamic programming and allows the use of gaps, which are calculated as an additional cost (gap penalty). The Needleman-Wunsch algorithm is a well-proven method in bioinformatics for gene sequence alignment. It uses standardized matrices to represent the mutation probability of each nucleotide base. Analogously, a matrix would be needed that includes the probability of each possible sound change. In practice, these transitions’ probabilities (i.e., the probability of a sound change) must be estimated, since a database with corresponding probabilities does not exist so far. Its creation proves difficult not only due to the lack of data from largely unexplored languages but also because of the way dialects are handled. A sound change can extend across dialectical boundaries — sometimes in a divergent form. The transition probability for a dialect is therefore not independent from the occurrence of the sound change of the neighboring dialect.

In order to determine the probability of a sound change between two sound classes, A and B, the PMI score (*pointwise mutual information*) is defined. PMI (Church and Hanks 1990) is a well-established measure for recognizing sound correspondences in CHL and belongs to the so-called *language-independent* approaches. In contrast to *language-specific* approaches, such as the LexStat algorithm in List (2012a), language-independent methods do not determine the recurring correspondences between the languages in question. While language-specific methods show better results, they do not do so for smaller datasets (cf. List 2014a). The PMI score is calculated as a binary logarithm of the number of co-occurrences of A and B divided by the theoretical number of co-occurrences when both are statistically independent ($p(A) \times p(B)$). Applied to sound changes with sound classes A and B, this yields the formula:

$$PMI(A,B) = log \frac{likelihood\ that\ A\ and\ B\ participate\ in\ a\ sound\ change}{likelihood\ of\ A\ \times likelihood\ of\ B}$$

A PMI score of 0.0 means statistical independence of A and B, which, however, will not occur in practice, since the sound to which a sound passes over is by no means random. The transitions are therefore expected to be positive or negative (e.g., for a transition from consonant to vowel).

Table 9.A. Results of the Needleman-Wunsch algorithm for the German concepts KATZE ‘cat’ and HAUSTIER ‘pet’ after ten iterations. The opening gap penalty was set to 5.5 and the samples were not IPA-transcribed. Etymologically related word parts are marked in italics.

Word Pairs KATZE ‘cat’		**Similarity Score**
Leo*pard*	Ge*pard*	99.95795
Katze	*Kate*r	22.67308
Katze	*Kätz*chen	0.23413
Tiger	Löwe	-30.91325
Kätzchen	Sphinx	-68.26321
Manx	Sphinx	-69.48069
Kater	Löwe	-90.05029
Katze	Löwe	-90.05039

Word Pairs HAUSTIER ‘pet’		**Similarity Score**
Hund	*Hünd*in	68.61254
Huhn	Hund	32.64941
Ochse	Mops	2.79437
Schwein	Hündin	-1.58880
Ross	Mops	-4.04401
Ruhn	Kuh	-10.20042
Kran	Katze	-15.80698
Manx	Kran	-17.35576

Instead of estimating the probabilities of the sound changes, the transition probabilities from the paradigmatic methods were used in this work. The logarithmized amount of the calculated PMI score from the previous iteration round is added as weight for the current round. Up to ten iterations were performed. Positive similarity scores indicate etymological relatedness (Jäger 2015:12756). The result for the semantic field of German KATZE 'cat' is illustrated in Tab. 9.A. Accordingly, the word pairs *Katze-Kater* 'tom-cat', *Leopard* 'leopard'-*Gepard* 'cheetah', and *Katze-Kätzchen* 'kitten' can be considered as potential "internal cognates." Etymologically, this is correct, since *Leopard* and *Gepard* represent borrowed compounds or univerbization with Latin *pardus* 'panther' (cf. Duden 2020:315). Compounds of this kind, as well as productive derivations such as *Kätzchen*, are true positives, but they do not form the prime target of this method. On the other hand, other concepts such as HAUSTIER 'pet' from Tab. 9.A show false cognates in the positive similarity scores. While the best-scoring word pair *Hund* 'dog'-*Hündin* 'she-dog' goes back to the same root, the other word pairs are based on chance. Especially short words are affected by this.

9.3 Evaluation

Since the method has a high error rate in determining the internal cognates, an evaluation of the sound correspondences will be omitted at this point and an evaluation of the internal cognate pairs will be carried out instead. This is to estimate how many of the identified pairs represent partial cognates. Partial cognates are all pairs that go back to a common origin or have cognate compound members (e.g., bio*logy*:phono*logy*). Ignored are cases in which a simplex lexeme is unchanged as a composite stem (e.g., *dog*:she-*dog*) or only affixes are cognate (e.g., *de*scribe:*de*compose). For the evaluation, 57 random concepts were determined, for which a maximum of 100 word pairs per concept were manually checked for relatedness. The resulting gold standard contained 4,693 word pairs with 1,445 different words. The evaluation achieves a precision of 0.1471 and a recall of 0.3526. The F-score is 0.2075. The best results were achieved by the concepts GEBÄCK 'biscuit' (with one true positive and one false negative; F-score 0.6667) and NAME 'name' (with six true positives and eight false negatives), which were the only concepts to achieve a precision of 1.0.

Table 9.B. Proportion of etymologically related words among the word pairs identified by different methods as cognates. For the semantic and paradigmatic approach, 4,693 word pairs were examined for this, for the derivational approach 4,363 word pairs.

	Precision	Recall	F-Score
Semantic approach	0.147059	0.352564	0.207547
Paradigmatic approach	0.999881	-	-
Derivational approach	0.999720	-	-

In order to have a comparison, the number of actual internal cognates of the paradigmatic and derivational approaches was added to Tab. 9.B. For this purpose, 4,693 paradigms and 4,363 derivations of a lexeme were checked for suppletion. The proportion of paradigmatic suppletion is limited to a few word stems (e.g., *stehen* ‘to stand’, *gehen* ‘to go’, or *-mann*:*-leute* ‘-man’), which reappear in several derivations. The percentage share of non-cognate derivations is higher. Most of these are words that were erroneously listed in the Wiktionary as word formations or were not recognized as comments by the automation. Nevertheless, both methods show a proportion of over 99.9% of actual internal cognates.

9.4 Discussion

To summarize the findings, unlike external cognates, the methods of automatic cognate recognition are less suitable for the identification of internal cognates. While there are usually lexemes of two languages with the same meaning that are compared, IR works with lexemes from a completely semantic field. This increases the risk of a random equation, especially for wordforms with few sounds. Moreover, these methods tend to be designed for simplex lexemes. Alignment assumes sound correspondences between strings with the possibility of a gap. Comparing simplex lexemes with derivations (such as *Hund-Hündin* or *Katze-Kätzchen*) or two compounds having a same component stem (*Leopard-Gepard*) is a new challenge for cognate recognition.

If it is possible to identify and eliminate the unwanted pairs (e.g., the random word pairs) in an adequate way, the identification of the wanted pairs seems to be possible, but their number is too small compared to the number of expected false word pairs (which includes loan words such as *cat* and *kitten*). This makes the determination of sound change hardly possible. Similarly, not all intentional pairs could be identified automatically by this method (e.g., *Hass* 'hate' ~ *Hetze* 'baiting'), since the ontology did not provide any relation between these concepts. The method is, therefore, most suitable for identifying new cognates from a list of potential candidates and for identifying unproductive morphemes, as illustrated in the following section.

9.5 Application Case: Proto-Indo-European

9.5.1 Preprocessing

9.5.1.1 Input Data

As application case, the semantic method is tested with Proto-Indo-European data. Proto-languages offer the advantage that their vocabulary is already listed as roots in dictionaries and accordingly do not require lemmatization. The most used dictionaries of Proto-Indo-European are LIV (*Lexikon indogermanischer Verben*) for verbs, NIL (*Nomina im Indogermanischen Lexikon*, Wodtko et al. 2008) for nouns and adjectives, and *Lexikon der indogermanischen Partikeln und Pronominalstämme* (Dunkel 2014) for particles. In order to not overcomplicate the semantic relations with different word types, this sample study will confine itself to the largest dictionary LIV and thus to verbal roots.

Unlike its predecessors, such as Pokorny (1959), LIV ignores morphological anomalies between roots or considers them synchronically unproductive (cf. LIV 2001:3–4). Root extensions or the so-called *schwebeablaut* cannot be explained synchronically and are word formation methods of the Pre-Proto-Indo-European stage. The extended roots **u̯remb-* 'to turn' and **u̯reng-* 'to twist, to bend' are therefore separate entries in the dictionary, as are the *schwebe-ablauting* roots **ḱeu̯H-* 'to throw, to push' and **ḱu̯eH-* 'to throw'. The function of root extensions is considered largely unknown (Fortson 2010:78–79), although various proposals have been put forward (e.g., Pisani 1974:93). For the outcome of this study, this means that a large proportion of identified cognates will belong to this kind of internal cognates and, therefore, the semantic method will be illustrated by the internal reconstruction of ancient and unproductive morphs instead of sound changes.

9.5.1.2 Ontology

Since LIV specifies all translations in German, the German ontology GermaNet 9.0 is used for this study. GermaNet is the largest German semantic network with over

144,000 synsets (May 2020) and provides six relations for verbs (hypernyms, hyponyms, causation, active entailment, passive entailment, and association). Using German ontology avoids translation errors via third languages.[62] However, mistranslation can occur as well when a word sense of a polysemous German verb does not correspond to the meaning of the Proto-Indo-European verb. For instance, the German verb *finden* may mean both 'to find' and 'to have an opinion'. Sorting out the wrong word senses is not possible without great effort. In practice, this means that more co-hyponyms and hypernyms are considered than intended.

9.5.1.3 Matching Proto-Indo-European Verbs with Synsets

Since not all meanings given in LIV have their own synset in GermaNet, some meanings have to be assigned separately. Verbal phrases that are usually missing in GermaNet can be assigned to a synset if it is present in the associated paraphrase in GermaNet. Other verb phrases are assigned to single words, ignoring auxiliary verbs or particles (e.g., *hilfreich sein* 'to be helpful' → *hilfreich* 'helpful', *ins Wasser eindringen* 'to enter water' → *Wasser* 'water' + *eindringen* 'to enter'). Verbs with a prefix are mapped to the simplex lexeme if the lemma does not exist in GermaNet. 16 verbs could not be mapped to a synset in this way and were mapped manually to a similar word (e.g., *liebgewinnen* 'to grow fond of' → *lieben* 'to love'). Overall, 74.46% of the given meaning in LIV were present in the ontology.

Many roots of LIV contain indeterminable phonemes, which are mostly marked with capital letters or brackets. In these cases, all possible forms of the root are considered in the process (e.g., **k(*$^{u̯}$*)enH-* 'to fall into joy' → **k*$^{u̯}$*enh*$_1$-, **k*$^{u̯}$*enh*$_2$-, **k*$^{u̯}$*enh*$_3$-, **kenh*$_1$-, …).

9.5.2 Weighted Sequence Alignment

The Needleman-Wunsch algorithm requires starting probabilities that describe the transition probability of a sound change. In addition to the lack of such data, the exact pronunciation of proto-sounds, especially the laryngeals, is not sure. Provisionally, the German transitional probabilities of the paradigmatic approach are used as starting probabilities in this case. Proto-Indo-European sounds unknown in German get the probabilities from phonetically similar sounds (e.g., [x,h] → laryngeals, voiced plosives → *mediae aspiratae*, [v] → [w], velars → palatals). Sounds that did not occur in Proto-Indo-European are deleted.

[62] For example, German *gießen* corresponds to PIE **ǵ*^{h}*eu̯-* but German *gießen* = English *cast* does not correspond to the PIE verb.

For each root, all relations of the synset with the same meaning available in GermaNet are considered. A potential source of error lies in the translation step. For a German word, there are usually several word senses, some of which do not apply to the Proto-Indo-European verb. However, determining the correct word sense is not straightforward. The problem is the retranslation of the German hypernym into Proto-Indo-European. In the procedure, all word senses and translation possibilities are considered. For a verb just like **reu̯H* 'to tear open', this results in 16 different hypernyms (cp. Fig. 9.2).

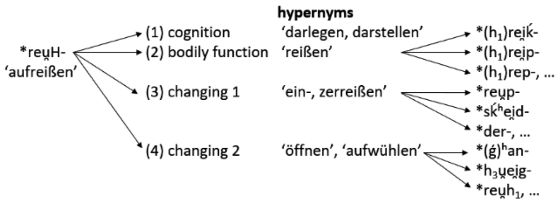

Figure 9.2: Sample of an assignment of a dictionary entry to an ontology concept. For the verb **reu̯H* 'to tear open', there are four word senses and, thus, four hypernyms. Since the hypernyms can be translated differently into Proto-Indo-European, 16 different hypernyms arise for this verb.

Among the hypernyms, an internal cognate of the hyponym is searched for in the next step. For doing this, the alignment must necessarily be based on concretized root forms, which means for a root such as **k(u̯)enH* 'to fall into joy' every possible wordform must be formed (e.g., $*k^{u̯}enh_3$, $*kenh_1$). For all possible forms of a hyponym and hypernym, a Needleman-Wunsch algorithm with a gap probability of 5.5 is performed. As described in the previous section, up to ten iterations are tested and for each round, the amount of PMI scores of the previous iteration was multiplied as weight.

App. C.21 shows the aggregate PMI score for **reu̯H* 'to tear open' with all potential cognates. It can be seen that the scores decrease with increasing round. This is due to the fact that there are hardly any new recurrences of sound correspondences among the internal cognates. In the first round, the cognate pair **reu̯H-* und $*h_3reu̯k$- 'to dig out' has a high aggregate PMI score because the starting probabilities for the transition from k-h_3 and h_3-*Ø* are still relatively high (h_3 was equated with the German sounds [x] and [h]). These correspondences could not be confirmed in the Proto-Indo-European data and became less important with each pass.

Table 9.C. The Proto-Indo-European word pairs with the highest aggregate PMI scores after five iterations. The open gap penalty was set to 5.5. Verb pairs with root extensions (RE) are marked in the third column. For more details, see App. C.22.

Aggr. PMI	Word Pair	RE
20.33019	*ˀdu̯er- ‘to run’ : *u̯er- ‘to run’	
	*ǵʰu̯er- ‘to walk wrily’ : *u̯er- ‘to run’	
	*u̯er- ‘to run’ : *u̯erǵ- ‘to act, to make’	
	*u̯erǵ- ‘to act, to make’: *u̯er- ‘to hinder, to ward’	
	*u̯erǵʰ- ‘to tie’ : *u̯er- ‘to run’	
14.83019	*dreu̯- ‘to run’ : *sreu̯- ‘to flow, to stream’	
	*ˀdu̯er- ‘to run’ : *u̯erǵ- ‘to act, to make’	
	*ˀdu̯er- ‘to run’ : *ǵʰu̯er- ‘to walk wrily’	
	*ǵʰu̯er- ‘to walk wrily’ : *u̯erǵʰ- ‘to tie’	
	*Hu̯er- ‘to enclose, to put in’ : *u̯erǵ- ‘to enclose, to lock’	
	*reu̯H- ‘to tear open’ : *reu̯p- ‘to break, to rip’	X
	*reu̯s- ‘to grub’ : *dreu̯- ‘to run’	
	*su̯er- ‘to hurt’ : *u̯ers- ‘to wipe away’	
	*tu̯er- ‘to actuate, to move sth.’ : *u̯erǵ- ‘to act, to make’	
	*tu̯erH- ‘to seize’ : *u̯er- ‘to run’	
	*u̯erǵ- ‘to act, to make’: *u̯erǵʰ- ‘to tie’	
	*u̯erǵ- ‘to act, to make’: *u̯erp- ‘to turn back and forth’	
	*u̯erǵ- ‘to act, to make’: *ǵʰu̯er- ‘to walk wrily’	
	*u̯erǵʰ- ‘to tie’: *ˀdu̯er- ‘to run’	
	*u̯erǵʰ- ‘to tie’: *u̯erǵ- ‘to enclose, to lock’	
	*u̯erǵʰ- ‘to tie’: *Hwer- ‘to enclose, to insert’	
	*u̯ergᵘ̯- ‘to throw’ : *u̯erǵ- ‘to act, to make’	
	*u̯ergᵘ̯- ‘to throw’ : *u̯erǵ- ‘to act, to make’	
	*u̯erp- ‘to turn back and forth’: *u̯ert- ‘to turn around’	X
	*u̯ert- ‘to turn around’: *u̯ers- ‘to wipe away’	
11.74426	*u̯reyḱ- ‘to turn, to enwind’ : *ˀu̯reyt- ‘to turn, to twist’	X
9.33019	*bʰreu̯s- ‘to break’ : *preu̯- ‘to jump’	
	*bʰreu̯s- ‘to break’ : *reu̯p- ‘to break, to rip’	
	*dreu̯- ‘to run’ : *ˀdʰreu̯b- ‘to drop’	
	*dʰreu̯bʰ- ‘to break into pieces’ : *preu̯- ‘to jump’	
	*dʰreu̯bʰ- ‘to break into pieces’ : *reu̯p- ‘to break, to rip’	

	*_ǵ^{h}u̯er_- ‘to walk wrily’ : *_tu̯erH_- ‘to seize’	
	*_h_2u̯erg_- ‘to turn (around)’ : *_u̯er_- ‘to hinder, to ward’	
	*_reu̯h_1_- ‘to open’ : *_reu̯H_- ‘to tear open’	X
	*_tu̯erH_- ‘to seize’ : *_u̯erǵ_- ‘to act, to make’	
	*_tu̯erH_- ‘to seize’ : *$^?$_du̯er_- ‘to run’	
	*_u̯erǵ^h_- ‘to tie’ : *_tu̯erH_- ‘to seize’	
	*_u̯erh_1_- ‘to say’ : *_u̯erǵ_- ‘to act, to make’	
	*_u̯reng_- ‘to twist, to turn’ : *$^?$_u̯rep_- ‘to decline’	X

Tab. 9.C shows the word pairs with the highest aggregate PMI after five rounds. The complete list with all word pairs (with an aggregate PMI score > 0.0) is given in App. C.22. The dominance of roots with [r] and [w] can be explained by the German starting probabilities and the high recurrence of these sounds in the Indo-European vocabulary. A complete evaluation seems difficult since no gold standard exists on Proto-Indo-European root extensions. The word pair *_u̯erǵ_- ‘to act, to make’ and *_u̯er_- ‘to run’ could formally be a root extension pair, but the semantic proximity does not necessarily speak for it. However, a development of meaning ‘to run → to effect, to make’ is conceivable. Without these cases of doubt, 16 of the positive pairs can be recognized as root extensions (precision = 0.3019). Among the false pairs, there are also pairs that could be potential cognates from both purely phonetic and semantic similarity (e.g., *_u̯er_- ‘to run’ : *$^?$_du̯er_- ‘to run’ or *_reu̯h_1_- ‘to open’ : *_b^{h}reu̯H_- ‘to break open’). A large proportion of incorrect pairs can be attributed to a semantic relation that is too distant. Pairs such as *$^?$_d^{h}reu̯b_- ‘to drop’ : *_dreu̯_- ‘to run’ have the relation DROP IS THE HYPONYM OF RUN in GermaNet.

Besides, there are cases of unrecognized root extensions, such as *_k^{W}eyt_- ‘to notice, to recognize’ : *_k^{W}ey_- ‘to perceive, to notice’. Its aggregate PMI score of -15.7766 is very low, which can be explained by the starting probabilities. These come from German data, where neither the diphthong /ei̯/ nor a postvocalic /i̯/ exists. The category of false negatives also includes _schwebeablaut_ pairs (e.g., * _$h_2$leks_- ‘to defend, to protect’ : *_$h_2$elk_- ‘to ward’). Of these, only one case was recognized (*_u̯reg_- ‘to follow a track’ : *_u̯er_- ‘to run’). Here, metathesis poses a methodological challenge.

9.5.3 Internal Reconstruction of Morphology

According to the procedure's theory, all word pairs with positive aggregate PMI scores are internal cognates. In the next step, these can be used as input for the procedure of Sect. 8.1. Since it is less likely to reconstruct sound changes on the basis of low precision values, the semantic method is used in this section for the internal reconstruction of unproductive morphemes. For instance, from the alternation *Ø-ḱ/k* of the words **teh*$_2$- 'to melt, to thaw' and **teh*$_2$*ḱ/k*- 'to melt (intr.)', an "intransitive morpheme" *ḱ/k* could hypothetically be determined. In individual cases, this determination is uncertain, since a variety of arguments can be offered against the hypothesis (e.g., incorrect reconstruction of the phonetic form or meaning, change of meaning or an incorrect relation).

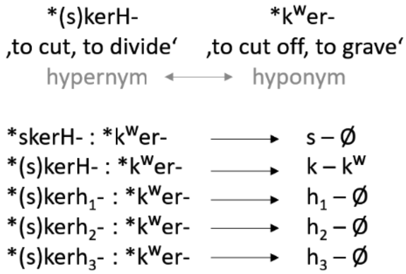

Figure 9.3: Sample of the determination of sound correspondences. The identified sound correspondences are assigned to the relation of the corresponding word pair (e.g., *k:k*^{w} → hypernym-hyponym).

In order to obtain sufficient potential cognates for an alternation, the number of iterations is limited to a few rounds. GermaNet offers six verbal relations, but only the relations of hypernymy and hyponymy occur sufficiently and frequently among the cognate pairs to make a statistically significant statement. From the identified cognates, all possible sound correspondences are extracted (cf. Fig. 9.3). Once enough data are available for the process, false sound correspondences should lose probability due to recurrence. These alternations are assigned to the respective relation. Since there is a mirroring between hyponymy and hypernymy, the word pairs occur twice (cf. Tab. 9.D).

The aim of the next step is to record whether the alternation-relation correlation is random or statistically significant. Such a test can be performed with the chi-square four-fold test (cf. Sect. 8.1.3.2). Fig 9.8 lists the alternation-relation pairs with the corresponding chi-square values. A statistical significance ($p > 0.95$) results for three different alternations:

- *k* is a feature for hyponymy
- *d^h* is a feature for hypernymy
- *i̯* is a feature for hypernymy

App. C.23 lists the cognate pairs with the lowest Levenshtein distances. It is rather unlikely that these three elements served in Pre-Proto-Indo-European to form the abstract relations of hypernymy or hyponymy. It is more likely that the concrete relations remain hidden by changes in meaning and reconstruction inaccuracy. The correct relation is not reconstructable even by a manual review of the cognate pairs.

Table 9.D. Result of the four-fold test with a degree of freedom set to 1 after one iteration. All pairs with $p > 0.95$ are considered significant. The number of hypernyms and hyponyms of *b*:Ø and Ø:*b* is too small for a significance test.

Quantil	Alternation	Relation	Number of Hypernyms	Number of Hyponyms	p-Value
7.86255	Ø : k	hypernym	28	53	0.995
7.76925	k : Ø	hyponym	54	29	0.9947
~~7.02744~~	~~b : Ø~~	~~hyponym~~	~~7~~	~~0~~	~~0.992~~
~~7.00212~~	~~Ø : b~~	~~hypernym~~	~~0~~	~~7~~	~~0.9919~~
4.85755	Ø : d^h	hyponym	39	22	0.9725
4.79505	d^h : Ø	hypernym	22	39	0.9715
4.70722	Ø : i̯	hyponym	40	23	0.97
4.64468	i̯ : Ø	hypernym	23	40	0.9689

Among the cognate pairs of *k*-HYPONYMS, only **d^heh_1-* ‘to put, to make’ : **d^heh_1k-* ‘to make, to produce’ forms a classical example of root extension, but the origin or function of the *k*-element is unclear (cf. LIV 139). Attempts are frequently made to explain this extension with Pre-Proto-Indo-European sound changes (e.g., Sturtevant 1940, Olsen 2010). A large proportion of word pairs show *k* rather in *anlaut* position (e.g., **ḱu̯ei̯t-* : **ku̯ei̯t-*, **(s)per-* : **(s)ker-*). To deduce a Pre-Proto-Indo-European preverb *k* from this, however, is not warrantable.[63] The study only indicates a significant accumulation of *k* in hyponyms, which can also be caused by other

[63] A *k*-prefix has been discussed by Poultney (1963:407–408).

factors. For instance, **kuei̯t-* is a variant of **ḱuei̯t-* 'to shine brightly', which only occurs in Balto-Slavic, where **ḱuei̯t-* is also attested (cf. LIV 375). Possibly, a Balto-Slavic sound development is the reason for the proto-form **kuei̯t-*. Another explanation for this could be the fact that words attested in few languages get more concrete reconstructed meanings than words attested in many languages. The interpretation of the d^h-HYPERNYMS is clearer. *$^?$*skerd*^{h}- 'to cut, to prick (?)' is a root extension of **(s)ker-* 'to clip, to scratch, to cut off'. The reason why d^h appears here as a significant feature for hypernyms also has to do with the abstract meaning reconstruction compared to the other root extensions of **(s)ker-*. The significance is thus influenced by the semantic reconstruction. The *i̯*-HYPERNYMS has already been recognized as a Pre-Proto-Indo-European infix. According to Ackermann (2018), the *i̯*-infix functioned to form intransitives and states or inchoatives from verbs. Even though only a few of the identified examples are good examples of the infix, the dominance of roots with the structure *CViC* among hypernyms argues for a separate semantics of the infix.

9.5.4 Discussion

Even though this study has shown that a Pre-Proto-Indo-European morpheme can by no means be directly inferred from statistical significance and that other possibilities exist, it has also shown that it is capable of potentially identifying a pre-language's morpheme from correlations. In the end, the interpretation of the candidates is to be left to the linguist. Especially IR on proto-languages always runs the risk of merely reproducing the ideas of the reconstructing linguist and the reconstructed meanings of proto-languages are often imprecise. The use of fitting ontologies that are more strongly based on relations or diachronic relation developments leads to the expectation of more specific relations in the future. To reduce the errors caused by the translation steps, the ontology should ideally be created for the language in question. Another optimization opportunity lies in the selection of the starting probabilities for the Needleman-Wunsch algorithm. By using cross-linguistic data, more adequate results may be achieved that do not depend on the occurrence of a sound change in a concrete language.

10. Phonotactic Method

A purely distributional method of IR has been repeatedly addressed in the literature without concretizing the method in a detailed algorithm. Rather, the feasibility of the method has been questioned. Nevertheless, in this thesis, a focus is placed on the very approach. The arguments for it are impressive: the procedure does not need any morphological knowledge. As explained in Chapter 6, synchronic morphophonemic alternations often lead to no new knowledge beyond a diachronic projection. In reconstructed proto-languages, alternations and morphology are constructed by the linguist, thus the linguist himself determines the outcome of an IR. Moreover, the number of alternations in most languages of the world is very small, and its reconstruction potential is therefore quickly exhausted. For the automation of IR, the morphophonemic method likewise implies an additional effort in resource acquisition since paradigms and derivations of a language have to be provided for this purpose. In this case, the phonotactic approach only requires a corpus or word list, ideally transcribed in IPA. This also makes it interesting for proto-languages or largely unknowable or unexplored languages. The aim of the following chapters is the question of the possibilities and feasibility of the distributional methods in principle, as well as the elaboration of the conditions that are necessary for a suitable implementation of the methods. The methods are tested for an easily evaluable language (German) and a proto-language (Proto-Indo-European) since the methods set different conditions for these types of language. In the literature, methods based on the phonotagms of a language are combined with methods based on the frequency of sounds. Formally, these are different approaches that require their own procedures. The two types also differ in that they attempt to reconstruct different sound changes: conditional (phonotactic) and unconditional sound change (frequency-based approaches). In this thesis, I distinguish between a phonotactic, a distinctive, and a gap approach.

The term “phonotactic approach” is understood in this thesis as the internal reconstruction that attempts to derive historical sound change via the distribution of phones and phonemes (i.e., the phonotactics) in the words of a language. Consequently, it starts from the basic hypothesis that synchronic phonotactics is the result of diachronic sound change. In Sect. 4.2.2., the problems and concerns raised in the literature have already been mentioned. The aim of this chapter is, in addition to the automation itself, to test the validity of the method. This includes:

- is the basic hypothesis valid (i.e., do phonotagms reflect sound change)?
- and can sound changes be inferred from phonotagms?

Only if the method reaches its limits, additional information should be added to the automated IR.

10.1 Theoretical Foundations

The integration of a sound law into the phonotactics of a language can be represented with a three-stage model. At this point, the German sound change *s > ʃ* / #_C (e.g., MHG *slange* → NHG *Schlange* 'snake') is given as a leading example. Its occurrence leads in the first stage (i.e., stage of a primary split) to a uni- or bilateral distribution: There is neither #*sC* nor #ʃV in the phonotactics (i.e., *bilateral* or *complementary*), or solely no sequence #sC (i.e., unilateral) since the phonetic sequence #ʃV already existed in the preliminary stage. In the second stage, the strict division is dissolved by loanwords, analogical levelling, or new word formations (e.g., German *slawisch* 'Slavonic'). However, at this stage, the distribution is still recognizable in a large part of the vocabulary (i.e., stage of unproductivity). In the final stage, the old sound change is obscured by younger sound changes and is no longer recognizable to the speaker (i.e., stage of secondary splits).

10.1.1 First Stage

In the first stage, the sound change is perceived as a synchronic phonological rule. The type of sound change has a direct influence on the phonotactics of the language in question. Especially the division between *merging* and *shifting* is of interest. In the case of shifting without condition, such as Proto-Germanic **w* > NHG [v] / _, the sound distribution in the corpus is hardly or not affected at all, so that its reconstruction by means of phonotactic methods is not to be expected. Shifting with condition leads to ALLOPHONES in the first stage. As an example, the development of the velar fricative to [ç] in German can be cited, which has led to the allophony of /ch/. This type of sound change best reflects the notion of sound change in the theoretical model of phonotactic reconstruction so that its reconstruction by this approach seems theoretically most appropriate.

More complex effects on phonotactics result from merging. Merging without condition results in a phoneme whose frequency in the corpus is the sum of the frequencies of the preceding phonemes in the previous stage. However, the increased frequency in the new corpus cannot be detected by the phonotactic method without external corpora. Merging with conditions shows similar results as shifting with conditions, although the term "allophony" is not unjustly avoided here. As an example, final devoicing leads to the fact that each final /d/ is pronounced as a voiceless [t]. This case leads to an absence of voiced obstruents in final positions in the corpus, while at the same time, voiceless obstruents appear disproportionately often in this position. It can therefore be assumed that a phonotactic reconstruction is at least partially possible for this type of sound change.

10.1.2 Second Stage

In the next phase, the synchronic sound change has become unproductive — one may speak of a "diachronic sound change." Loanwords, neologism, or analogically newly-formed words that no longer follow the old phonological rule, may indicate this transition in the corpus. An example is the already mentioned *s* > *ʃ* / #_C that is not productive in German anymore, as minimal pairs just as [stiːl] 'style' and [ʃtiːl] 'stalk' or [slam] 'slum' and [ʃlam] 'mud' illustrate. Since these pairs are only formed with loanwords or abbreviations and are limited to a small number, the old sound change is often still recognizable to speakers. For a computational approach, however, this phase means that a purely binary evaluation (i.e., the sound either occurs or does not occur in position *X*) cannot be purposeful but must be replaced by a relative result (i.e., the sound occurs in position *X comparatively* rarely or frequently).

10.1.3 Third Stage

In the last phase, the sound change is no longer recognizable for the speaker (e.g., because it is hidden by secondary sound changes). An example is the assimilation of nasals in Early Germanic. This was hidden in German by dissimilation (cf. Gothic *fi*m*f* and NHG *fü*n*f* 'five') or syncope (e.g., NHG *He*m*d* 'shirt'), or secondary assimilation (cf. NHG *sanft* 'soft' in common speech [samft]) distorts the actual phonetic diachrony. Whether and to what extent these sound changes can be reconstructed by the phonotactic approach is a question that will be discussed in this chapter.

10.2 Procedure of the Phonotactic Approach

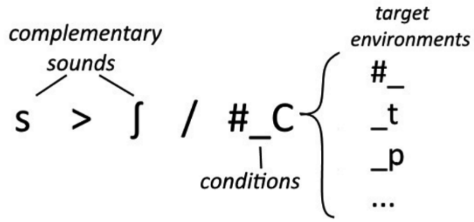

Figure 10.1: Sample of a sound change and explanation of the terminology used in this chapter.

The goal of the procedure is the reconstruction of recent sound changes from word lists or corpora. For this purpose, occurrence vectors are generated in accordance with the model of distributional semantics (Turney and Pantel 2010, Padó and Lapata 2007, Bullinaria and Levy 2007). The dimensions of the vectors represent the frequency of one or more neighboring sounds. In the next step, we are looking for vectors with complementary distribution (i.e., the vectors of complementary sounds) which show a maximum distance in the searched dimensions but only differ minimally in the others. As an example, the German phoneme /ch/ and its allophones may be mentioned: [ç] occurs after *i* and *e* — so it has a high value in these dimensions — while the values for the position after *a* and *o* tend towards zero. This relation is exactly reversed (i.e., *complementary*) for the other allophone [x]. Such contexts are relevant for the sound change (i.e., target environments). In the other environments (e.g., before [ə]) both complementary sounds behave similarly (cp. Fig. 10.1). For the computational approach, these steps have to be divided into three parts: preprocessing of the data, normalization and weighting, and determination of complementary sounds.

10.2.1 Preprocessing

10.2.1.1 Input Data

Text corpora and word lists are suitable as source data. By using word lists, a larger and evenly distributed vocabulary is used, which leads to the expectation of a better recording of regular alternations. CORPORA offer the advantage that the most frequent words contain relatively few loanwords, while word lists often contain a disproportionately large number of borrowed words. In this way, the proportion of borrowed phonotagms could be reduced. The disadvantage of corpora is word repetition, which can bias the relevance of phonotagms that occur in common words.

This problem can be circumvented by using a LIST OF WORDFORMS, but then again, there is the problem that there are a large number of inflectional morphemes, making certain phonotagms more prevalent (e.g., English phonotagms with *s* due to the verbal ending of the 3rd sg. and the regular plural ending of nouns). Similarly, derivational morphemes (e.g., prefixes and suffixes) occur more frequently in a LEMMA LIST. The fourth alternative is the use of MORPHEME LISTS, but again, it must be taken into account that sound changes derivable at morpheme and word boundaries can no longer be identified. The German word [bɪndən] 'to bind' appears as [bɪnd] in a morpheme list, which means that final devoicing cannot be reconstructed. For the evaluation, all data will be extracted from a word list taken from the online dictionary Wiktionary, which provides IPA transcriptions.

10.2.1.1 Occurrence Vectors

For each different sound in the corpus or word list, one *feature vector* is created. Each dimension of the vector represents a feature (i.e., a sound environment) and its value is the frequency of the sound in question in that environment. The features may be defined in many ways. The feature patterns *before_X* and *after_X* seem to be particularly suitable to detect sound changes, in which *X* stands for each sound and initial or final word position. In a language with 25 different phones, a sound vector consists of 25+1 features per feature pattern. With two feature patterns, there are 52 features for each sound. Taking the untranscribed sentence "Lorem ipsum dolor sit amet, consectetur adipisici elit" as an example, the values of the features *before_t* and *after_t* for the vector of *e* are 2 and 1, respectively. The vector for *e* with only these two dimensions may be defined as

$$\vec{e} = \begin{pmatrix} before\ t: 2 \\ after\ t: 1 \end{pmatrix}$$

With the two mentioned feature patterns *before_X* and *after_X* only sound change conditions can be recognized that follow these patterns. A sound change with the condition *X_X* as with *s* > *ʃ* / #_C is not recognized or only its individual components #_ and _C. The use of more complex patterns could therefore reconstruct more accurate conditions, but this could also be subject to bias due to sparse data problems.

10.2.2 Normalization and Weighting

To avoid bias in the final result, it is necessary to replace the absolute frequency values with normalized and weighted values. Since the method has strong similarities to the methods of text mining, the usual measure (Beel et al. 2016) for normalization and weighting, *tf.idf* measure (Spärk Jones 1972), is used as a starting point.

$$tf.idf = \frac{freq_{ij}}{\sum_k freq_{ik}} \times log \frac{|D|}{\{d|t \in d\}}$$

10.2.2.1 Term Frequency

The first factor $\frac{freq_{ij}}{\sum_k freq_{ik}}$ represents the term frequency (tf_{ij}) and is used to locally weight and normalize the absolute frequency ($freq_{ij}$) of a word j in document i. Transferred to our case, the term frequency corresponds to the frequency of the feature (i.e., the corresponding context, such as *before_r* and *after_i*). In the sample sentence above, these are 2 and 1, respectively. Document i corresponds to the sound in question, which is *e* in the sample. For normalization, the absolute term frequency of a word j is divided by the number of each word k in document i, resulting in a relative value. Besides, other variants exist for normalizing the absolute term frequency, such as dividing by the sum of the term frequency of word j in each document d ($\text{tf}_{2ij} := \frac{freq_{ij}}{\sum_d freq_{dj}}$). The latter is denoted as tf_{2ij} in App. D.5.

All relevant phonetic environments for the German sound [ç] (i.e., all front vowels and all preceding consonants) occur as frequently as expected so that five of the eight most frequent phonetic contexts belong to the conditions for [ç] and thus to the target environments. The use of tf_{2ij} shows a better result: six of the eight most relevant phonetic contexts are target environments. This can be explained by the fact that phonetic contexts that occur frequently, in general, receive a larger denominator and consequently a lower tf_{2ij} value.

10.2.2.2 Inverse Document Frequency

In text mining, the inverse document frequency $\log\frac{|D|}{\{d|t \in d\}}$ is used for the global weighing of the term. If the term occurs in all documents, which is the case for articles or the copula, the *idf*-value is log(1) and accordingly receives a low weight, while the weight of relevant content words is high. Transferred to sound environments, the low total number of documents (|D|) corresponds to the count of phones, and the number of documents containing the term in question ($\{d|t \in d\}$) corresponds to the number of sound environments that occurs with the phone. Due to the low number of phones and the fact that most sound environments do appear next to most phones, *idf* is not a useful weight for the phonotactic approach (cf. App. D.6). It is therefore reasonable to switch to other document frequencies that have an underlying threshold of more than 0.0 ($\{d|term \in d\} = \{d \,|freq_{term\ in\ d} > 0\}$). Thus, two additional limits are applied in App. D.6. *Idf-h*$_p$ uses as denominator the number of phones that have a higher value for the context than the respective sound *p* (in App. D.6, this is [ç]):

$$idfh_p := log \frac{|D|}{\{d \,|freq_{term\ in\ d} > freq_{term\ in\ p}\}}$$

Idf$_m$, on the other hand, determines the arithmetic mean and uses this as the threshold value:

$$idf_m := log \frac{|D|}{\{d \,|freq_{term\ in\ d} > \frac{\sum_{i=1}^{n=|D|} freq_{term\ in\ i}}{|D|}\}}$$

Rare sound environments, such as *before_ʒ*, receive a higher weight with all measures, which is only partially useful for this purpose. In App. D.6, a relevant weighing of the target environments is only shown by the *idf-h*$_p$ value, in *which the target environments after_œ, after_ɪ* and *after_ʏ* receive the highest weighting value.

10.2.2.3 Relevance Measure

From the previous two sections, a new measure of relevance for phonetic contexts emerges. Since phonetic contexts may be absent from the corpus in smaller datasets, distorting zero values can occur. To overcome this sparse-data problem, Laplace smoothing (Lidstone 1920) is applied, in which corpus frequencies are increased by one (cf. Carstensen et al. 2010:155–156). Thus, the following formula is obtained as a relevance measure:

$$rel(j) := \frac{freq_{ij} + 1}{\sum_d freq_{dj} + N} \times log \frac{|D|}{\{d \,|freq_{term\ in\ d} > freq_{term\ in\ p}\}}$$

Tab. 10.A only shows minimal changes in the ranking of sound environments between *tf*$_2$ and the relevance measure. However, the weights lead to a stronger separation of the values of the target environments from the values of irrelevant sound contexts. The procedure then seeks to determine the most relevant context among all sounds (e.g., *ɪç*: after_*ɪ* for the sound *ç*). For the determined sound, the sound change's conditions and the possible complementary sound are searched in the next step.

Table 10.A. The ten most relevant conditions of [ç] in a German corpus of 1,000 sentences with the relevance measure rel(j) and its factors. Rel(j) is calculated from the values in the columns 2 and 4. The real conditions of [ç] (i.e., after front vowels and consonants) are highlighted in italics. Without smoothing.

tf$_2$		**idf$_h$**		**rel(j)**	
after_ɪ	0.20688	before_ʒ	3.66356	*after_ɪ*	0.75792
after_œ	0.20522	*after_œ*	3.66356	*after_œ*	0.75183
after_ʏ	0.17112	*after_ɪ*	3.66356	*after_ʏ*	0.62690
after_ɛ	0.05937	*after_ʏ*	3.66356	*after_ɛ*	0.13520
before_t	0.05732	*after_ɛ*	2.27727	before_t	0.13053
after_ɛː	0.04653	*after_ʒ*	2.27727	*after_ɛː*	0.08709
before_#	0.03173	before_t	2.27727	before_#	0.05450
after_ʁ	0.01603	*after_ɛː*	1.87180	before_ə	0.01909
before_ə	0.01508	before_ŋ	1.87180	*after_l*	0.01283
after_l	0.01440	before_x	1.71765	*after_ʁ*	0.01239

10.2.3 Determination of the Conditions of Sound Change

In the previous section, it was shown how "relevant" sound environments can be determined and measured for a sound. In this context, the term "relevance of a sound environment" refers to the value of a sound environment for a sound, which measures the relative frequency of that environment compared to other sounds. Consequently, the sound occurs significantly more often in these contexts, so that a random distribution cannot be assumed here. This value does not provide a hint of a sound change by itself. However, since after a conditional sound change a complementary relationship has been formed between the sounds involved to a certain degree, which depends, among other things, on whether it was a shifting or merging process. However, the relevance values can provide an indication of the conditions of the preceding sound change to a certain degree. As in the morphophonemic method, two possible relevance values are offered: a percentage measure and a difference measure, in which for each sound environment k, the relevance values for the two sounds i and j are subtracted, and the values are sorted numerically. The difference measure calculates the vector of each condition $cond_{ij}$ as follows:

$$cond_{ij} = \begin{pmatrix} rel_{k1i} - rel_{k1j} \\ rel_{k2i} - rel_{k2j} \\ \cdots \\ rel_{kni} - rel_{knj} \end{pmatrix}$$

Table 10.B. Results of subtracting the relevance values of [ç] from [x], [ʃ] from [s], and [d] form [t]. For this calculation, the German data from CELEX were used. The mid-range values have been omitted. The correct target environments of the sound pairs are highlighted in italics.

ç - x		ʃ - s		d - t	
after_ɪ	0.75792	*before_p*	1.09715	before_ʒ	3.33051
after_œ	0.75183	*before_l*	0.05566	before_e:	1.52728
after_ʏ	0.62690	*after_#*	0.04072	before_i:	0.96981
after_ɛ	0.13520	before_œ	0.02116	before_a	0.51488
before_t	0.10644	before_ø:	0.01530	after_#	0.51034
after_ɛ:	0.08709	after_œ	0.01107	before_ɛ	0.46251
…	…	…	…	…	…
before_j	-0.01242	after_ɪ	-0.29379	after_x	-0.80816
after_u:	-0.04648	after_y:	-0.33824	before_ɛ:	-0.81799
after_a	-0.04677	after_t	-0.42681	before_s	-0.95338
after_ʊ	-0.05999	after_a	-0.49286	after_ç	-1.22274
after_ɔ	-0.08416	after_ɛ	-0.73709	after_ʃ	-1.23096
after_a:	-0.39427	before_u:	-1.56847	after_ɛ:	-1.24253

Tab. 10.B shows three examples. The first example maps the synchronic allophony of German [ç] and [x]. the sound environments with the highest values correspond to the conditions of the subtrahend, the lowest values to the "counter-conditions" of [ç] (i.e., the conditions of [x]). A restriction to five sound environments results in the rules from the first column Tab. 10.B:

- /ch/ → [ç] / { ɪ, œ, y, ɛ }_, _t
- /ch/ → [x] / { a:, ɔ, ʊ, a, u:}__

These rules already approximate the actual distribution rule of /ch/ (/x/ → [ç] / [+front]_, [+consonant]_). The second example, [ʃ] and [s], targets the sound change *s > ʃ* / #_C. This is a conditional merger of *s* with *ʃ*, so there are no "counter-conditions." The first three conditions (*before_p, before_l*, and *after_#*) correctly reflect this sound change, but this procedure cannot distinguish between #_C and #_/_C, which may be avoided by the additional use of "trigram contexts." Moreover, the complete abstinence of chronologically close sound change (e.g., MHG *hirz* →

Hirsch 'deer') is striking. This is due to the small number of example words reflecting these sound changes. Since the method is based on a corpus-based and coarse-grained procedure, it is not suitable for identifying "small sound changes" and rare conditions.

The third example in Tab. 10.B represents the synchronic final devoicing of /d/. In contrast to allophony, *d* and *t* coincide in final positions (cf. *Rad* [raːt] 'wheel' and *Rat* [raːt] 'council'). Fig 10.5 shows that the target environment *before_#* is ranked only ninth among the conditions for *t* (i.e., the ninth-last condition for *d*). Many of the "relevant" contexts for *t* such as *before_s* and *after_ʃ* may reflect older assimilations. Other contexts such as *after_œ* and *after_ɛː* do not reflect any sound change and are to be evaluated as "false positives." Their appearance is likely to have morphological causes. Since *t* occurs frequently in both inflectional morphology and derivations, the sound appears significantly more often in "non-target environments." This problem could be remedied by a list of roots instead of a corpus, but this suffers from other disadvantages (e.g., final devoicing would not be identifiable).

The three examples show the limits but also the possibilities, which result from the phonotactic method. An open question is that of the threshold value. A possible relative value for the first example (e.g., ±5% of the maximum value 0.758 ≈ 0.038 or the first five contexts) is not suitable for sound changes without "counter-conditions" as in examples 2 and 3 in Tab. 10.B. Absolute threshold values are unsuitable, which is already illustrated by the highest value 3.33 in the third example that is not assigned to any target environment. Example 3 is also an example of the partial invalidity of the basic hypothesis. As explained in Sect. 10.2, the method assumes that statistical anomalies can be attributed to sound change. However, this assumption cannot often be confirmed — at least not directly from the phonotactic distribution. The comparison of the frequencies of different difference values for the allophonic sound pair *ç-x* and the irrelevant sound pair *k-ʊ* illustrates that for the "irrelevant" sound pair *k-ʊ* a comparable distribution of contexts is given as for *ç-x*. From this, it can be concluded that no statement about a sound change can be made from a significant deviation in the phonotactic distribution itself. Whether the distribution reflects a sound change or not must therefore be decided via the complementarity of the sounds in question.

10.2.4 Determination of the Complementary Sounds

The determination of complementary sounds represents one of the greater challenges of the phonotactic approach. While the morphophonemic method uses alternations as basis for the determination of sound pairs, this must be done on the basis of phonotagms, which, however, cannot be done immediately. A comparable problem can be found in distributional semantics for the identification of antonyms, co-hyponyms, or other semantic relations. For this purpose, a large part uses pattern-based methods (Schwartz et al. 2015) or additional sources of information, such as thesauri (Yih 2012 and Ono 2015) or ontologies. These, however, are not suitable for the phonological domain. Despite the widespread assumption that purely distributional models are not suitable for the identification of antonyms (cf. Scheible et al. 2013:439), several approaches have already been proposed in research based on the attempt to determine suitable features that can serve as identification (e.g., Scheible et al. 2013 with word classes or Nguyen et al. 2016 with weighting of synonym and antonym classes). Meanwhile, a direct transfer of these methods to reconstruction methods is not possible or solely in a highly modified form. In this thesis, several approaches are tested including purely phonotactic, phonetic, and typological approaches. Since most sound changes take place between "similar" sounds, these approaches try to identify complementary sounds on the basis of a concept of similarity. This similarity is differently defined by the approaches.

10.2.4.1 Phonotactical Approaches

If one is willing to follow a purely phonotactic approach, "similar sounds" can only be understood as phonotactically similar sounds. Phonotactic similarity of sound vectors can be formally determined by simple distance measures, such as cosine similarity or Euclidean distance. But in most cases, it is precisely the "dissimilarity" that is sought (cf. Sect. 10.1). A simplified approach to determining complementary sounds of a sound s_1 is to separately evaluate the distance to both the outlying and all the other environments. Contexts that occur unusually frequently with the sound are indicative of a sound merging, making this context absent or low for the complementary sound. However, the absence of occurrences is not unusual for many sounds. For instance, vowels generally occur infrequently after front vowels, giving rise to a large number of potential complementary sounds in this context. In addition to determining the outlying contexts, therefore, the distance of a potential sound from the sound s_1 serves as a factor. Theoretically, the sound s_1 shows similar values in other contexts, while vowels here diverge strongly phonotactically. However, the equation of target environments and statistical outliers turn out to be questionable in practice, as has already been illustrated in Sect. 10.2.3. The "non-target environments" of *d-t* have the largest outlier values, while the highest values of *ç-x* are not outliers in the mathematical sense.

Table 10.C. The most relevant bigram contexts of German *ʃ* according to the *featSelect* formula. Setting the threshold to 0.05, *_p*, *_t*, and *_v* would be the target environment of a sound change ? > *ʃ* / _{*p, t, v*}.

featSelect values for the conditions of [ʃ]			
1.06715	before_p	0.01968	after_ə
0.07824	before_t	0.01629	before_øː
0.07065	before_v	0.01135	before_ɛː
0.04922	before_l	0.00736	before_aː
0.03790	after_#	0.00522	after_ɪ
0.02837	after_œ	0.00466	before_ɛ
0.02017	before_œ	0.00410	after_ʁ

To select the "outlier" environments (*feature selection; featSelect*), the features (i.e., the sound environments) of a sound are weighed by how strongly the environment appears in similar sound vectors. To determine the similarity of two sounds, cosine similarity is applied, which is multiplied by the feature value. Note that cosine similarity here only measures the phonotactic similarity of the sounds, not their phonetic similarity. The average value gives the relevance of feature (i.e., environment) *i* to sound *j* (with *n* is the number of sounds):[64]

$$feature_{ij} - \frac{\sum_{k \in S}^{n} feature_{ik} \times cosSim(\vec{j}, \vec{k})}{n}$$

As an example, the feature-selection values for the sound *ʃ* are listed in Tab. 10.C. The relevant features reflect the conditions for the sound change *s* > *ʃ* / #_C. The environments with positive feature-selection values occur more frequently with the sound than expected. The environments that exceed a threshold value θ (e.g., 0.05) can then be evaluated as relevant. To determine the complementary sound, the values (i.e., similarity measure and the values of the selected features) are compared.

[64] This formula is also suitable for optimizing the procedure of determining the conditions. However, it does not provide negative values and, thus, no "counter-conditions" for allophones.

Tab. 10.D shows the complementary sounds for *ʃ* (historical sound change: *s* > *ʃ* / #_C), *t* (*d* > *t* / _#), *k* (*g* > *k* / _#), and *ç* (*x* > *ç* / _V[+front]) determined in this way. It can be seen that the correct complementary sound appears among the relevant sounds but cannot sufficiently distinguish from other candidates. They belong to the first type of *sound change with few contexts*. The determination of relevant contexts is often too specific for individual sounds to be derived from the phonotactic distributions alone. The second type is *sound changes with competing contexts* that cannot be determined as such because other sound changes make the determination of the complementary sound more difficult. Thus, one of the relevant conditions for the German sound *k* is ɛ_s reflecting the sound change *x* > *k* / _s. Another type of sound change is found in column 4 (sound *x*) whose complementary sound [ç] appears in the lower relevance range. The reason for this is that it is a *sound change with many contexts*. The contexts V_ and C_ make up the bulk of the contexts, so there are few "irrelevant" contexts. The consequence is that the use of a similarity measure here leads to the opposite end — the sounds are phonotactically too dissimilar.

Table 10.D. The most probable complementary sounds for the German sounds [ʃ, t, k, x], calculated with the formula $f(c) = \left(1 - \frac{\sum_i^n feature_{ic}}{n}\right) \times cosSim(s, c)$. Bi- and trigram contexts and a threshold of 0.5 were taken into account. The correct complementary sound is highlighted in italics.

ʃ		t		k		x	
p͡f	0.06924	*d*	0.08503	f	0.06602	χ	0.04682
s	0.04310	m	0.07866	*g*	0.06590	ɐ̯	0.02191
k	0.03939	p	0.07822	p	0.06536	f	0.02092
f	0.03878	l	0.07601	l	0.06041	l	0.01749
t͡s	0.03247	n	0.07099	n	0.05854	n	0.01580

10.2.4.2 Phonetic Approaches

In this section, further linguistic resources are tested for the determination of the complementary sounds. For this purpose, the occurrence of all other sounds in the significantly relevant sound environments of the target sound are considered (i.e., *complementary distribution*), which can be defined as the sum of the *tf-idf* values for these contexts. The sound that fulfils the lowest sum and an additional linguistic condition may be determined as the complementary sound. As "additional linguistic condition," four several principles are tested.

According to Chafe (1959:479), a sound pair can be determined to be "similar" if they share at least one common distinctive feature (CHAFE'S SIMILARITY PRINCIPLE). The major classes *vowel* and *consonant* are ignored. Accordingly, the sound with "complementary distribution" with one common distinctive feature is chosen as complementary sound.

The second principle assumes that sound change occurs between articulatorily similar sounds (HOENIGSWALD'S SIMILARITY PRINCIPLE, cf. Hoenigswald 1965:74). Accordingly, the complementary sound is the sound with complementary distribution that is most similar to the sound in question. A purely formal similarity measurement via the number of shared distinctive features leads to [t] (*voiceless, alveolar, plosive*) being equally to both [d] (a*lveolar, plosive*) and [ʔ] (*voiceless, plosive*), while [t] only shares one feature with [t͡s]. To account for this, the same place of articulation receives a value of 2 and any distance from it receives a subtraction of 0.5 (see App. D.10).

Another principle works by considering the distinctive features in the conditions of the sound change (common CO-OCCURRENT FEATURES IN SOUNDS AND CONDITIONS). The distinctive features that are dominant in the conditions are regarded as the features that the complementary sound had lost in the assimilation process. Accordingly, the complementary sound is most similar to the sound in question without these specific distinctive features. The features *vowel* and *consonant*, which are most dominant here, are ignored. Similarly, the vowel features *front* and *back* must be equated with the consonantal equivalents *palatal* and *velar* in order to map an assimilation process.

The fourth principle works with DOLGOPOLSKY CLASSES. According to Dolgopolsky (1964), sound changes occur most frequently between a few sounds, which he typologized in their own sound change classes. The so-called Dolgopolsky classes are the traditionally most used sound change classes in CHL (e.g., in Baxter and Ramer 2000) and originate from the time of classical lexicostatistics. The complementary sound for the sound in question must belong to the same class. For each sound, however, only a few sound changes are possible, while assimilations are not taken into account. More recent alternatives, such as the sound classes of ASJP (Brown et al. 2008) or List (2010:44), extend the classes. List (2010) includes 28 sound classes, including vowels or tones and five other consonant classes. However, classes for assimilation are missing here as well, and many classes contain only a single or no sound for many languages (e.g., [j] in as the only approximant in German).

Table 10.E. List of relevant environments (bigrams and trigrams) of the German sound [x]. The *featSelection* threshold was set to 0.2.

featSelect	Relevant context	featSelect	Relevant context
3.579796	oː_d	0.794885	ˈaʊ̯_t͡s
1.192353	ˈaː_f	0.794822	aː_ɪ
1.590735	ˈaː_b	0.680272	ˈaː_ʁ
1.590509	ˈuː_ʃ	0.262687	ˈʊ_ʃ
0.953937	ˈuː_h	0.230076	ˈʊ_#

Example of the Application | In the following example, the list of relevant features for the German sound [x] and a *featSelection* threshold of at least 0.2 is given in Tab. 10.E. Based exclusively on these features, there is a high number of sounds that do not occur in this context (f(c) = 0.0). Among these, vowels dominate in particular, which usually occur rarely in postvocalic positions in German. According to the four methods mentioned in the last section, the sound can be determined as complementary sounds of [x] that fulfilled the respective condition.

- Chafe's principle: *ç, v, ʒ* (fricative), and *ŋ̍* (velar)
- Hoenigswald's principle: *k* (velar, voiceless)
- Dolgopolsky's principle: *ç* and *d͡ʒ* (class 4)
- Condition principle: *ç, v,* and *ʒ*[65]

The comparison of the four principles shows that the correct complementary sound [ç] does appear among the candidates, but in no case could it be clearly determined as such. Tab. 10.G shows the result for other German alternating sounds. Hoenigswald's, Dolgopolsky's, and the condition principle have more often similar results since both are based on Hoenigswald's principle that sound change occurs between similar sounds. However, this principle narrows down the possible candidates much more, so that especially sound changes between dissimilar sounds like [ʁ] and [ɐ] are excluded in advance and cannot be identified. This is particularly pronounced in the Dolgopolsky classes. The sound *t* only occurs in the second Dolgopolsky class

[65] The dominant features of the conditions (cf. Tab. 10.C) are *velar/back* and *voiceless*. The feature *velar/back* occurs frequently in the conditions because the sound [a] was listed as an "unrounded open front vowel" in this evaluation.

(dental obstruents without sibilants), which consists of only two German sounds ([t] and [d]). Consequently, the method could only determine the sound [d] as a complementary sound, while no prediction at all could be made for the vowel due to the lack of vowel classes. The correct prediction of the condition principle depends strongly on the correct prediction of the context's dominant features. It is not uncommon to determine the features *voiced* and *voiceless* as the dominant feature, since these form the most common features among phonemes. If the feature is another, the method, again, either fails altogether (e.g., for [ʁ] and [ɐ]) or provides a great number of possibilities (e.g., *ç, v, ʒ* for *x* since we are looking for a non-velar fricative). However, this method has the best results of all phonetic methods. The method nevertheless seems to find suitable complementary sounds, if there is an underlying assimilation process. It should be noted that the results with the assumptions of Chafe and Hoenigswald may be worse with more phonemes considered.

On the other hand, the method that does not require a maximum similarity of the sounds is much freer in its prediction, but it is also more difficult to determine the correct complementary sound. In principle, there are too many candidates that occur in complementary distribution as well: in German, the dental vibrant [r], just like [d], does not occur in the word-final position (in this context, /r/ becomes [ɐ]), just as the syllabic nasal consonants rarely appear in the relevant contexts of voiceless plosives.

Threshold and Relative Chronology | The *featSelection* value reflects the dominance of a condition given the frequency of this condition in other sounds. Conditions with values > 0.0 occur significantly more often with the sound than with other sounds, which means that a relevant context must be above a threshold > 0.0. If this threshold is too close to zero, irrelevant contexts can also be included in the evaluation. If it is too high, the number of candidates increases, since fewer contexts are classified as relevant and thus more sounds reach the required frequency deficit in the "relevant" contexts.

Discussion | The methods based on articulatory similarity of sound pairs are able to predict the correct complementary sound in most cases, especially when the sound change is an assimilation process. However, they prove to be incorrect when the actual complementary sound is phonetically very different from the sound in question and no assimilation (e.g., for ʁ-ɐ) is present. Since this knowledge cannot be derived purely phonotactically, the methods can only provide information about the complementary sound under ideal conditions.

10.2.4.3 Subtypological Approach

If for a language A_1 with the randomly generated vocabulary W = {*word₁, word₂, ... wordₙ*}, one simulates a PHONEMIC SPLIT leading to bilateral allophony (such as *x* > *ç* / e,i_), this implies for the subsequent language A_2 the splitting of the old phoneme into two new sound vectors. These are positioned close to the zero values compared to the previous sound vector of A_1. The pre-sound (in the example, *x*) takes the value zero for the conditions (i.e., e_ and i_), while the new sound *ç* shows no evidence for the counter-conditions (e.g., a_). In non-normalized vectors, the same relative values then apply in all the other conditions (e.g., _e) so the sum of the new complementary sounds is identic to the pre-sound: $\overrightarrow{a_2} + \overrightarrow{b_2} = \overrightarrow{c_1}$. Fig. 10.2 illustrates this evolution in a vector space for the pre-sound vector $\overrightarrow{\chi_1}$ (marked in bold), which evolved into the sounds $\overrightarrow{\chi_2}$ and $\overrightarrow{ç_1}$ in language A_2.

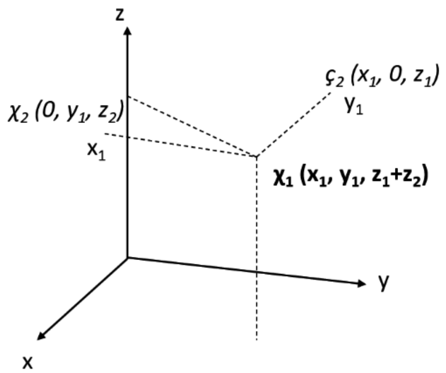

Figure 10.2: Illustration of a sound change in the vector space. The pre-sound is marked in bold, the new sounds in italics. The dimensions x, y, and z represent the relevant environments for χ_2 (= y), $ç_2$ (= x), and irrelevant environments (z).

For a CONDITIONED MERGER, as in the case of final devoicing, this rule does not apply, since one of the two sounds from language A_2 has two original sounds in A_1 — consequently, *a* can go back to *d* or *t*. The sum of $\overrightarrow{a_2}$ and $\overrightarrow{b_2}$ thus still coincides with the sum of the same sounds from language A_1:

$$\overrightarrow{a_2} + \overrightarrow{b_2} = \overrightarrow{a_1} + \overrightarrow{b_1}$$

This rule can also be applied to phonemic splits, if one of the two vectors of language A_1 (namely the sound newly created in A_2) is given the value zero, while in the case of the unconditioned merger, the zero vector is to be used for $\overrightarrow{a_2}$ or $\overrightarrow{b_2}$. In order to be able to determine the complementary sound for sound *a* with this procedure, a transformation of the formula and a determination of the sound vector which comes closest to *b* is sufficient.

To illustrate this, a conditioned merger was simulated in a German corpus, where [k] was changed to [t] before [ɛ] and [ɪ], as well as a phonemic split with newly introduced [c] in the same environment in another corpus. Then, the occurrence vector[66] of [k] was added to the occurrence vector of each sound of the language, and the addition of the same sound vectors was repeated with the original corpus. According to the formula, the two added vectors for language A_1 and A_2 must be as similar as possible, showing a low Euclidean distance. In the simulated data, the complementary sounds show a Euclidean distance of 0.0 and thus stands out clearly from the other sound vectors (Tab. 10.F).

Table 10.F. The table shows the potential complementary sounds for *k* in the first test case (conditioned merger, left columns) and second case (phonemic split, right columns). ED indicates the Euclidean distance of k_1+b_1 and k_2+b_2. This is zero for the respective correct complementary sounds (*t* and *c*, respectively) and highest for the sound in question (*k*). Non-normalized data with the conditions *before_sound* and *after_sound* were used.

ED	Complementary sounds (case 1)	ED	Complementary sounds (case 2)
0	*t*	*0*	*c*
83.9404551	ʦ, ç, d, uː, l, b, z, p, j, f, h, œ, v, ʊ, x, ʧ, χ, g, ʔ, yː, øː, ʒ, ʤ	83.940455	ʦ, ç, d, uː, l, b, t, z, p, j, f, h, œ, v, ʊ, x, ʧ, χ, g, ʔ, yː, øː, ʒ, ʤ
83.9523674	m, ə, oː, ʃ, eː, ʏ	83.946411	m, ə, oː, ʃ, eː, ʏ
83.9880944	a, aː, n, ɔ, ɛː	83.964278	a, aː, n, ɔ, ɛː
84.0476056	iː	83.994047	iː
84.3682405	ŋ, ɐ, s	84.15462	ŋ, ɐ, s
84.5221864	r	84.231823	r
98.3361582	ɛ	91.422098	ɛ
105.337553	#	95.241798	#
118.659176	ɪ	127.645645	ɪ
167.88091	*k*	*167.88091*	*k*

[66] In this test, the pure occurrence vector without any normalization was used.

The test makes it clear that complementary sounds can be determined purely phonotactically if an older language stage is available as a reference corpus. For IR, this possibility is not given. Only (sub-)typological data from other languages can serve as reference material (cf. Greenberg 1978:78–79). In subtypological contexts, the comparative data come from phonotactically strongly similar languages. In most cases, this is only true for closely related languages. This bears the risk that instead of typological information, phylogenetic-comparative data are included in the internal reconstruction process.

To implement the subtypological approach, the source data was taken from the UCLA Phonetics Lab Archive[67]. This dataset provides 60,351 IPA-transcribed words from about 300 languages worldwide. Since the number of sounds differs significantly from language to language, all sounds in the subtypological corpus must be assigned to a corresponding sound from the language in question that is phonetically closest to it. At the same time, stress marks, tone letters, syllabic marks, diphthongs, gemination, and affricates are eliminated or resolved in the corpora, as the transcription of the data is highly inconsistent. To determine the most phonotactically similar languages, a vector is constructed for each language, each indicating the relative frequency of bigram phonotagms. Cosine similarity is used to determine the similarity of the languages. For German, the five most similar languages are found to be Dutch, Norwegian (Bokmål), Plautdietsch, Catalan, and Danish. In the test, the procedure was repeated with the five, ten, twenty, thirty, forty, and fifty most similar languages, as well as with all languages.

Fig. 10.3 shows the rank of the correct complementary sound for the German allophones *x*-*ç*, the final-devoicing plosives, and the sound pair *s*-*z*, which is partially complementarily distributed due to the sound change *s̠* > *z̠* in pre-vocalic position (see App. A1 Sect. *(Pre-)Old High German Sound Changes*). The best results were obtained by considering 20–30 phonotactically similar languages. The test did not confirm the assumption that choosing languages that are as genetically close would improve the results. For the poor performance of the sound pair *d*-*t*, a large number of sound changes of these sounds (e.g., Germ. **d* > OHG *t* and Germ. **t* > *t͡s, s*) certainly also play a role, because these changes caused an unusual distribution of German *t* and *d* in phonotactic terms. Despite the high ranking of the correct complementary sounds for other sound pairs, the complementary sound nevertheless often did not reach the first rank. As a sole method for reliably determining sound pairs, the subtypological method is, therefore, not sufficient.

[67] http://archive.phonetics.ucla/edu (accessed in December 2020)

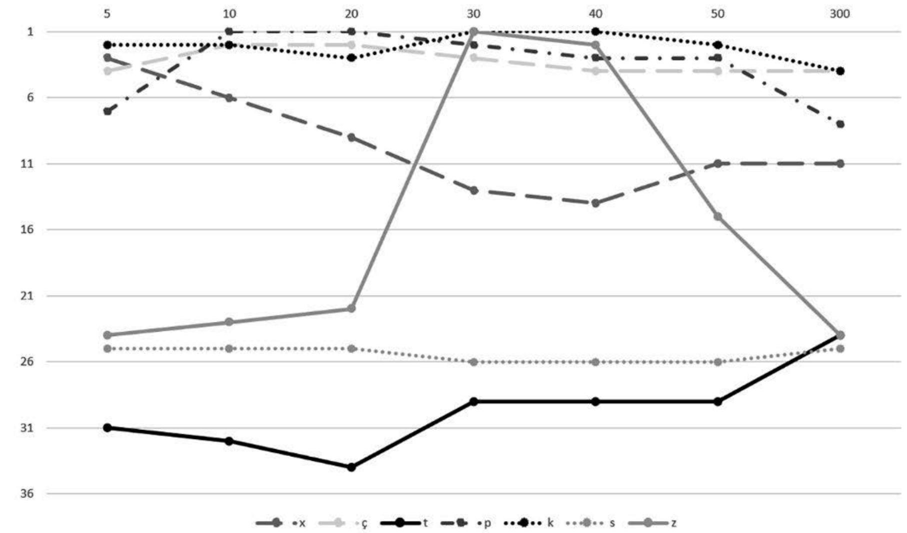

Figure 10.3: Ranking of the correct complementary sounds for the German sounds *x* (grey, dashed), *ç* (light grey, dashed), *t* (black), *p* (grey, dashed/dotted), *k* (black, dotted), *s* (light grey, dotted), and *z* (light grey) with *y* phonotactically similar languages as reference corpora. The data were normalized and only the relevant features (*featSelect* with threshold > 0.1) were used to determine the complementary sound.

10.2.4.4 Empirical-Probabilistic Approach

Another typological possibility is to determine the probability of a sound change from the sound changes documented worldwide and to choose as a complementary sound that sound which has the highest probability among the sounds with the lowest *featSelect* value. Such a database of sound-change probabilities does not exist; however, Brown et al. (2013) determined the frequency of recurrent sound correspondences in the ASJP database to provide an estimate of the occurrence of various cross-linguistic sound changes. Since no cognate list existed for many languages, they restricted themselves to a simplified procedure that considered all semantic word pairs as cognate pairs if they differed in only one sound position. In this way, all sound correspondences that occurred more than once in a language pair were considered recurrent. They only recorded correspondences (i.e., a statement about the change direction or the condition is not given). Furthermore, similar vowels are combined under one sound, diphthongs and long vowels are missing. Thus, the sounds from the language in question must be assigned to a sound or sound class from Brown et al. (2013). In order to determine the most probable complementary sound, a list of potential candidates is generated as in Sect. 2.4.2. Tab. 10.G shows that the empirical-probabilistic approach was able to predict the correct pre-sound in four out of nine cases, thus achieving a similarly good result as the conditions approach. However, it is also clear here that they increasingly lead to incorrect conclusions in the case of "unusual" sound changes such as r > ɐ.

Table 10.G. Determined complementary sounds for different German sounds using six different methods. Correct determinations are marked in italics. The threshold for a relevant context was set to 0.2. Syllabic sonorants were replaced in the source data. The sound *ç* is missing in the list of Brown et al. (2013). The subtypological approach used 30 reference languages.

Sound **Compl. sound**	**χ** **ç**	**ç** **x/χ**	**t** **d**	**p** **b**	**k** **g**	**r** **ɐ**	**ɐ** **ʁ**	**s** **z**	**z** **s**
Chafe	*ç*	h	x	x	x	j	øː	h	x
Hoenigswald	*ç*, h	h	t͡s	*b*	x	d	ə	t͡s	*s*
Dolgopolsky	*ç*	*x*	*d*	v	x	l	-	*z*	ʃ
Condition	*ç*, h	h	*d*	*b*	*g*	t͡s	ə	*z*	*s*
Subtypological	l	j	œ	j	*g*	ç	-	a	*s*
Empirical	x	-	*d*	*b*	*g*	l	œ, ɛː, ɛ	ʃ	*s*

10.3 Evaluation

The focus in this evaluation will be on finding relevant contexts. To evaluate the relevance of a potential sound change, the following formula from Sect. 2.4.1 is used:

$$f(c) = \left(1 - \frac{\sum_i^n feature_{ic}}{n}\right) \times cosSim(s,c)$$

10.3.1 Iterative or Non-Iterative?

Since the relevance of a context is relative to the frequencies of a context with other sounds, the contexts that are relevant may be "covered" by other relevant contexts. Thus, a context such as *_b* for *m* may be significantly frequent in relation to other consonants, while it is rare in comparison to vowels in that context. However, the theoretical model assumes an originally equal distribution of phones, which was only removed by sound change. In fact, vowels and consonants are distributed differently in syllable structures (i.e., *CV contrast*). The consequence of this is an overrepresentation of vowels and the word boundary (#) in the evaluation results (cf. App. D.7) since these sounds appear more frequent in the phonotagms. This effect may be dampened by an iteration procedure in which the best-rated sound change is removed after one step and its contexts are set to 0. This procedure also simulates RELATIVE CHRONOLOGY to some degree.

A comparison of the results from the iterative and non-iterative procedures shows a minimal difference between them (see App. D.8). Tendentially, the iterative procedure causes a slightly better ranking of the relevant contexts (e.g., *ɛ* / _ç and *ʊ* / _x). The dominance of the vowels and # cannot be significantly reduced in this way. As a result, many sound-context transpositions (see Sect. 10.5.2) remain in the final list. Iteration has a positive effect on the occurrence of low-frequency sounds (e.g., *ʒ, ʤ*, and *j*). Phonotagms such as # / _j, *iː* / ʒ_, or *n* / ʧ_ are dropped in many cases and the number of false positives could be reduced. The use of an iteration can therefore be recommended.

10.3.2 Optimal Source Data

The choice of the source data has a further impact on the results, as explained in Sect. 10.2.1.1. In this test, different data types are evaluated to check their suitability. The corpus used for the test was an annotated portion of the untranscribed TIGER corpus (Brants et al. 2004), which already contained the lemmatized form of each wordform.[68] In order to make a meaningful comparison, both a wordform list and a lemma list were extracted from this corpus.

In order to test the application with a morpheme list, a morphologically analyzed word list of the CELEX project (Burnage 1995) was used. The German morpheme analysis of the CELEX database recognized 9,550 different morphemes. This analysis breaks down the words into their morphemes and assigns them to a morpheme lemma (e.g., *Absichtserklärung* 'declaration of intent' → *ab+seh+s+er+klar+ung*). Faulty analyses (e.g., *abgeschieden+heit* 'seclusion' instead of *ab+ge+scheid+heit*) increased the number of morphemes in the final morpheme list.

App. D.9 compares the resulting phonotagms with an untranscribed corpus, a lemmatized corpus, a lemma list, and a morpheme list. The comparison shows only minor differences in the ranking of the phonotagms. Borrowed phonotagms (e.g., *n*

[68] For the test, I have used the file *tiger_release_aug07.corrected.16012013.conll09* [accessed in August 2019].

/ _ç or # / à_) appear slightly increased in the list of corpora because rare phonotagms have a smaller proportion in the larger mass and are thus considered more "relevant" in the evaluation. Phonotagms occurring in high-frequency words appear more often in the result list of the corpora, but their proportion is lower than expected (e.g., # / _w from words such as *war* and *welcher*, *e* / d_ from *der* or *den*). Similarly, prefixes appear more often in the word list (e.g., *e* / g_ from the prefix *ge*- and *e* / v_ from *ver*-). Lemmatization has only a very small influence, which is reflected in individual words (e.g., all German articles *der, die, das* were lemmatized to *der*, which is why *e* / d_ and *e* / _r appears higher in the ranking of the lemma list). Therefore, lemmatization does not seem to be mandatory.

Since the test data is untranscribed, the result list also shows SPELLING-CONDITIONAL RELEVANCIES. These include digraphs such as *qu* and *ch*, which can be considered relevant contexts (i.e., true positives) depending on the viewpoint. Just like true relevant phonotagms (e.g., #j), they occur with similar frequency in all lists. In the morpheme list, these true positives are better ranked, but direct comparison is not possible due to the different source data. The choice between the corpus and lemma list consequently makes little difference, although due to the higher ranking of borrowed phonotagms in the corpus list, there is a tendency towards word lists. The highest rankings of relevant contexts were obtained with the morpheme list.

10.3.3 Match Rate

The match rate indicates how many of the identified contexts can be assigned to a sound change or a phonological rule. The first 100 iterations of the German dataset are evaluated (see App. D.1). Since after one iteration several contexts can be identified, the algorithm has resulted in 102 different contexts for the evaluation data. Due to sound-context transpositions the condition and sound may be interchanged (e.g., *a* / _x instead of *x* / a_), what makes a sound change may appear repeatedly in the result list. In this case, the sound change is evaluated only once, and the other is categorized as duplicates. As duplicates are also considered those phonotagms which only reproduce the same sound change with another condition (marked in italics in Tab. 10.H).

In total, seven different sound changes and five phonological rules could be identified (see. App. D.2 and D.3). In addition, there were 14 doublets for sound changes and 28 for phonological rules. Dominant in the evaluation are phonetic contexts with low-frequency sounds. This leads to a larger number of the higher-ranked context being borrowed phonotagms (e.g., contexts with *ʒ*) and accounted for a total of 18.628% of the 102 contexts. About eight contexts can be attributed to affixes (e.g., the infinitive suffix *-en* and the prefixes *ver-* and *vor-*). The result after the first 15 iterations and all identified sound changes and phonological rules can be seen in Tab. 10.H.

Table 10.H. Selection of the result list of the phonotactic method, performed with German Wiktionary data (for more details, see App. D.1). The threshold was set to 0.5 and phonotagms were restricted to bigrams. The column "FeatSel." indicates the *featSelection* value of the best phonotagm of the iteration. Phonological rules and sound changes (PR/SC) which occur more than once (i.e., doublets) are marked in italics. False Positives that reflect morphemes have the type *FP*. The gold standards can be found in App. A.1 and A.2.

It.	FeatSel.	Phonot.	Type	Gold standard 1	Gold standard 2
1	2.45394	# / _h	PR/SC	x → h / _{-$,:}	-
2	2.38725	ɪ / _ç	PR/SC	x → ç / {-V[-front]}	ç → x / V[-front]
3	2.05555	# / n_	FP		
4	*1.95308*	*ɛ / _ç*	*PR/SC*	*x → ç / {-V[-front]}*	*ç → x / V[-front]*
5	1.86849	# / _j	PR/SC	ie [ɪə̯ˀ] → je / #_	
6	1.79784	n / ə_	FP		
7	2.17728	# / ə_	FP		
8	*1.77546*	*a / _x*	*PR/SC*	*x → ç / {-V[-front]}*	*ç → x / V[-front]*
9	1.72371	# / _ʃ	PR/SC	s̺ → ʃ / X_t	
10	1.67526	# / _b	FP		
…					
17	1.46614	a / ʔ_	PR/SC		Ø → [ʔ] / (_V…)$_\omega$
28	1.57318	k / ŋ_	PR/SC		n → ŋ / [-continuant, -sonorant, +velar]_
33	1.29511	# / ɐ_	PR/SC	ʁ → ɐ / _$	r → ɐ / .C_0_C_0. *or* ʀ → ɐ / .C_0_C_0. *or* ʁ → ɐ / .C_0_C_0.
38	1.25199	# / _z	PR/SC	z̺̊ → z / X_V and #_V	
54	1.10846	n / _d	PR/SC	t → d̥ / n_	

76	1.13882	# / s_	PR/SC		z → s / _[-sonorant]$_0$.

For the calculation of the evaluation measures, the doublets and free variants are removed, as these complicate the calculation of the recall. In addition, a "precision with doublets" is introduced that categorizes the doublets as "true positives." Tab. 10.I shows the evaluation measures for the best F-score, the best precision with doublets and the measures after 100 iterations. The best F-score was achieved after five iterations for phonological rules, and after 55 iterations for sound changes. However, this difference results from the different sizes of the gold standards and thus from the recall values. For both, the majority of the correctly identified rules can be found within the first 30 iterations. Due to the high proportion of doublets (i.e., one sound pair with different conditions), the precision with doublets is significantly better, reaching a hit rate of 80% for the sound changes after 5 iterations. Surprisingly, the scores after the first iterations perform better for sound changes than for synchronic phonological rules. This is caused by phonotactically strongly constrained sounds such as *h* or *j*, which only occur at the initial position and can be attributed to two historical sound changes. As can be seen from Fig. 10.4, the proportion of doublets is greater for phonological rules, which can be explained by their synchronicity.

Table 10.I. Evaluation measures of the best F-score, the best precision with doublets and the measures after 100 iterations.

		Phonological Rules	**Sound Changes**
Best F-Score	precision	0.19048	0.18605
	recall	0.15385	0.03704
	F-score	0.17021 (after 5 iterations)	0.06178 (after 55 iterations)
	precision with doublets	0.42105 (after 25 iterations)	0.8 (after 5 iterations)
After 100 iterations	precision	0.03165	0.09091
	recall	0.19231	0.03704
	F-score	0.05435	0.05263
	precision with doublets	0.32353	0.20588

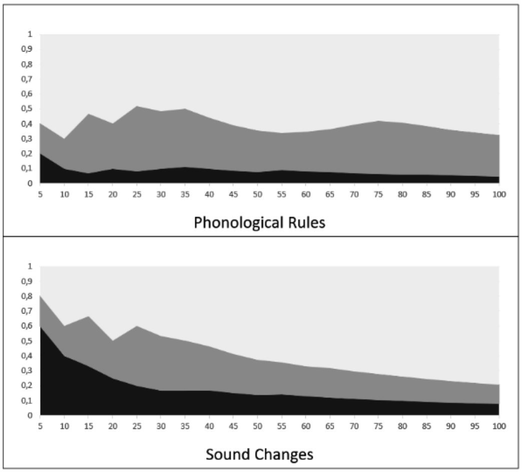

Figure 10.4: The diagrams show the percentage of different identified sound changes and phonological rules (PR/SC) after x iterations (black part). The grey part is the percentage of doublets; the light grey part the false positives.

10.3.4 Evaluation of Complementary Sounds and Sound Change Direction

The problematic nature of determining the complementary sound has already been pointed out in Sect 10.2.4. The majority of the identified sound changes and phonological rules are in transposed form (i.e., condition and resulting sound are interchanged). Since it is not possible to determine the correct complementary sound with wrong resulting sounds, only non-transposed sound changes and phonological rules after 500 iterations are used for the evaluation. App. D.4 shows the identified complementary sounds of all nine detected phonological rules and sound changes. Most of the sounds determined first by the approaches do not match the historical or synchronic complementary sound. The evaluation shows the inability of the phonotactical method to determine the complementary sounds appropriately. Although the correct sounds are not infrequently found among the top-ranked sounds, it was only rarely possible to determine them unambiguously.

Due to the fact that the phonotactic method does not determine sound pairs but phonotagms (i.e., the resulting sound of the sound change and its condition), the evaluation for sound-change direction and conditions are summarized in this section.

For their evaluation, the rate of non-transposed forms of the identified sound changes and phonological rules may be calculated. Of the ten different sound changes and phonological rules identified after 100 iterations, none show the correct change direction.[69] It must therefore be summed up that the determination of the direction of a sound change or phonological rule is not possible with the phonotactic method.

10.4 Application Case: Proto-Indo-European

The phonotactical method aims rather at the internal reconstruction of proto-languages, whose suitability for this shall therefore be tested as far as possible at this point. Due to the difficulty of determining the pre-sound, I restrict myself to the question of the validity of the basic hypothesis (i.e., are sound changes reflected in the phonotagms?). For this purpose, the experiment is repeated with Proto-Indo-European data and the "most relevant" phonotagms are interpreted linguistically and historically as far as possible.

The proto-language of the Indo-European languages is certainly the most researched reconstructed language. Its pre-stage is also a frequent topic of papers and monographs, in which the phonotactic peculiarities have also been thematized and taken into account (cf. Chapter 5). It is therefore not to be expected to gain new insights into the phonotactics of the language. Rather, the focus is on comparing the automated results with those of a human IR and phonological rules. This means that instead of a quantitative evaluation, a qualitative evaluation is used due to the lack of a gold standard and the low number of Pre-Proto-Indo-European sound changes and phonological rules.

10.4.1 Source data

The source data comes from the *Lexikon Indogermanischer Verben* (LIV) containing 1,170 verbal roots.[70] Uncertain reconstructions (marked with a question mark in LIV) are treated just like other reconstructions. Phonetic uncertainties marked with

[69] This could lead to the consideration that the identified conditions and sounds should be interchanged in a secondary step to achieve higher precision for sound-change direction. But it should be noted here that the procedure always determines the rarer sound of a bigram as a condition. It is, therefore, to a certain extent a coincidence that all ten detected phonotagms were present in transposed form.

[70] In addition, the dictionary *Nomina im Indogermanischen Lexikon* (Wodtko et al. 2008) with 207 nominal roots was also taken into account. Since these entries are often not listed in root form, it was necessary to sort them out from the evaluation.

bracketed sounds are eliminated in favor of the wordform with the bracketed sound. Since abstract sound groups (e.g., H for any laryngeal) would result in additional phonemes, all wordforms containing upper-case letters are removed. The final word list contains 1,012 root morphemes in this way, with duplicate entries due to homonyms.

An overview of Proto-Indo-European phonological rules is provided by Fortson (2010:68–72). However, this is by no means complete and contains only twelve phonological rules with some rules of post-Proto-Indo-European date.

10.4.2 Qualitative Evaluation

Table 10.J. The result of the phonotactic approach with Proto-Indo-European data after 14 iterations. Verbal roots from LIV were used and the threshold was set to 0.5.

It.	featSelect	Phonotagm	It.	FeatSelect	Phonotagm
1	2.56710	# / g_	8	2.16746	e / ḱ^h_ h$_2$ / b_ s / _ḱ^h
2	2.49437	# / b_	9	2.09298	# / d_
3	2.49506	e / _y	10	2.05327	e / d_
4	2.43839	e / p^h_ s / _p^h	11	2.03152	# / _s
5	2.43540	# / ǵ_	12	1.89537	# / g^h_
6	2.22332	e / ǵ_	13	1.85999	# / _k^W
7	2.21201	e / _n	14	1.82265	# / ǵ^h_

Some of the phonotagms recognized in Tab. 10.J map phonological rules and phonotactic anomalies that have already been discussed in the literature. There are several phonotagms in the list that show a voiced plosive (i.e., *media*) at the root-final position (Iteration 1,2,5,9, and 14). 72.332% of the non-labialized *mediae* in the source data are found in the final position. This is particularly clear for *g*, which is located in that position in 42 out of 52 occurrences. Of these, *w* precedes *g* in ten cases. This can be attributed to the βουκόλος rule (*boukólos*; see Fortson 2010:70), according to which a labiovelar is delabialized in the environment of *u* or *w,* respectively. The relevance of *b#* is due to a similar imbalance and a well-known phenomenon (Johansson 1900). Due to the lack of initial *b* in the word list, the sound occurs significantly at the root-final position. Consequently, the unequal distribution of *b*, which is generally rarely represented in Proto-Indo-European (so-called *b*-gap, cf. Kümmel 2007:236), in syllables is shown here. The discrepancy may also have been caused by developments of single ancient languages.[71]

The relevance of *ey* (iteration 3) results from the fact that *y* is an allophone of /i/, which appears exclusively before or after a vowel. The phonotagm, therefore, corresponds to the phonological rule *i* > *y* / V. The semivowel *w*, which is an allophone of /u/ in a similar way, ranks below since it can also occur in root-initial position before consonants.

Another phonological rule can be seen with the phonotagms *s* / _$\acute{k}^h$ und *s* / _p^h (iteration 4 and 8). According to Siebs's law, any root-initial *mediae* and *mediae aspiratae* becomes voiceless when preceded by an *s* (Siebs 1907). Through this rule, the Proto-Indo-European *tenues aspiratae* (p^h, t^h, $\acute{k}^h$, and k^h) almost exclusively occur after *s* and, in addition, the number of initial *mediae* decreases. Since this rule concerns the initial sound (i.e., it occurs before the vowel), the procedure additionally found the irrelevant phonotagms *e* / $\acute{k}^h$_ und *e* / p^h_ (iteration 4 and 8). The function of the prefixed *s*- (the so-called *s-mobile*) is unclear but may well explain the accumulation of initial *s* in the word list (cf. iteration 11).

Two factors can be considered as the reason for the significance of the phonotagm *en* (iteration 7): On the one hand, the peculiarity of Indo-European roots of positioning non-syllabic sonorants only before and after vowels (Tichy 2009:37) often results in the clusters CRV or VRC (e.g., *enC*). On the other hand, there was an *n*-infix in Proto-Indo-European, which is inserted after the root vowel and served to form present tense forms. If this spreads secondarily in the paradigms of the daughter languages, a reconstruction of the nasal infix as such becomes more difficult and may be reconstructed as a root consonant.

The interpretation of the results shows that a large part of the results does not represent known historical sound changes or synchronic phonological rules. Phonological rules captured by the procedure include Siebs's law, the allophony rule of /i/, and in part the βουκόλος rule. However, other phonological rules such as the dental-plus-dental rule (App A.3) have not been possible using the phonotactic

[71] Many of the roots ending with *b* are exclusively found in Germanic and Slavonic and occasionally have phonetically similar equivalents ending with b^h in other languages (e.g., *d^hreb* 'to drop' and *d^hreb^h* 'to break, to crumble' or **(s)kreb* 'to scrape, to scratch' and *$s\acute{k}/kreyb^h$* 'to scratch'). A monolingual development can therefore not be completely excluded (examples of secondary Germanic *-p* are given by Lühr 1988:351–361).

method — at least in the approach presented here. Of the twelve phonological rules of Proto-Indo-European mentioned by Fortson (2010:69–72), only voicing assimilation was not detected, which could theoretically be extracted from the data. It should always be borne in mind when interpreting the results of proto-languages that reconstructions inevitably present an incomplete picture of the language's phonotactics. For instance, the sound *t* seems to occur more frequently before the second laryngeal than before other consonants. This may have been different in the historical proto-language. Since *th_2* is reflected in Vedic Sanskrit as aspirated *th* (cf. Meier-Brügger 2010:247), this phonetic sequence may be easier to reconstruct than other consonant-laryngeal phonotagms.

10.5 Discussion

To sum up, two serious obstacles to the phonotactic method are identified in the thesis that avoids guaranteeing a valid result. Firstly, this is the identification of "noisy" phonotagms (i.e., do phonotagms reflect sound changes and phonological rules?) and, secondly, the identification of the correct positions of the sounds.

10.5.1 Do Phonotagms Reflect Sound Changes or Phonological Rules?

In this thesis, "irrelevant phonotagms" are understood as phonotagms that have neither been formed by sound change nor by synchronic phonological rules. Since the phonotactic approach can easily determine "potential" phonotagms, the irrelevant sound sequences are a major reason for the unreliability of the results. Their identification is therefore of particular importance for future work in order to obtain validation of the method.

The number of noisy phonotagms depends strongly on the format of the source data, which have different drawbacks that need to be taken into account (Tab. 10.K). Recurrent morphemes imply an overrepresentation of different phonotagms, so MORPHEMES are most likely to be chosen as source data.

Table 10.K. Disadvantages of four different source data: text corpus, wordforms, lemma list, and morphemes.

Source data	Disadvantages
Text corpus	Inflectional, derivational morphemes, and word repetitions distort the frequencies
Wordform list	Inflectional and derivational morphemes appear in the result list
Lemma list	Derivational morphemes appear in the result list
Morpheme list	Single relevant contexts (e.g., final devoicing) cannot be identified

There are many contexts with low-frequency sounds (e.g., *ʒ, ʤ*, and *j*) in the German result list. This is, in principle, intended by the phonotactic approach, since such a deviation in distribution indicates a sound change. In a purely internal and phonotactic way, however, it cannot distinguish between borrowed and inherited words, so many of the noisy phonotagms represent BORROWED PHONOTAGMS such as ʤ_ (cf. App. D.1).

The second group of noisy phonotagms consists of UNIVERSAL PHONOTAGMS (i.e., phonotagms which are dominant for purely articulatory reasons). These include primarily the contrary positions of consonants and vowels in syllables (*CV contrast*) or the juxtaposition of two voiceless consonants. Their automatic identification seems to be feasible to a certain extent, but the boundary between universal and sound change-related phonotagms is blurred. Phonotactic co-occurrence of voiceless consonants may have resulted from assimilations.

In addition, other noisy phonotagms could be identified that are specifically related to the source data. In non-transcribed data sets, SPELLING-CONDITIONAL PHONOTAGMS occur, such as digraphs. In proto-languages, some phonotagms may be overrepresented. Since the RECONSTRUCTABILITY of a sound can vary depending on the daughter language, phonotactic unevenness inevitably arises in the reconstruction. The extent to which a relevant phonotagm can be traced back to a historical sound change or a reconstruction obstacle often cannot be answered with certainty.

Without an automated determination of the noisy phonotagms, a manual interpretation of the results is always necessary. However, such as step is quite time-consuming and to a certain extent also requires specialist knowledge. Without a corresponding insight into the source data, it is often not possible even for an expert of the language to determine the reason for the relevance of a phonotagm.

10.5.2 Is it Possible to Determine the Pre-Sound and the Conditions?

The transition of a phonotagm to a sound change or phonological rule is difficult even from a theoretical point of view. To formally describe a sound change, all three positions of the pattern X' > X / _Y must be known. However, a phonotagm *XY* does not provide a direct way to identify X'. Three different ways of identifying the pre-sound, which is mandatory for the reconstruction of the pre-form, were tested. In this test, approaches based on phonetic similarity or empirical-probabilistic performed best. However, these approaches proved to be flawed in dissimilation processes. Determining the pre-sound therefore remains an unsolved problem.

A second issue that complicates the determination of a phonetic change X' > X / _Y is the correct position of X and Y. A phonotagm *XY* occurs once as X / _Y, once as Y / X_ in the result list, with the less frequent sound appearing first as a condition in the final ranking. A phonetic sequence *ux* is thus first captured as *u* / _x, although the correct change is *x* > *ç* / _V[+front]. Consequently, incorrect SOUND-CONTEXT TRANSPOSITION occurs when the context is more frequent than the changing sound. This is indeed true for the majority of bigram sound changes, so that the phonotagm with incorrect context was usually identified first in the ranking. The repeated occurrence of a context in the list (e.g., _ç or ʔ_ in App. D.1) may indicate such a sound-context transposition.

11. Distinctive Method

In the previous chapter, the distributional method was automated using the phonotactic approach. As explained in Sect. 4.2.2, approaches involving the relative phonetic frequency, or the symmetry of the phoneme system have also been put forward in the literature. These have only been presented using a few examples, which are by no means readily generalizable. In order to do justice to this aspect of the distributional method, this chapter proposes an implementation based on the distinctiveness of phonemes, which is therefore referred to as the “distinctive approach.”

11.1 Theoretical Foundations

The concept of the phoneme represents one of the most important innovations of structuralist linguistics. It is defined as the smallest distinctive sound in a language *that distinguishes from another* and thus no longer focuses on the articulatory essence of a sound but on its position in the system. This ability to distinguish forms the basis of phonology and is traditionally made tangible with minimal pairs.

The IR method described below is based on the idea that sound changes influence the distinctiveness of phonemes and that these leave traces in the phoneme system. The term *distinctiveness* can be defined in two ways. ABSOLUTE DISTINCTIVENESS is the “distinctive value” of a phoneme in the phonetic system. A simple measure of this would be the absolute frequency *freq* of a sound *s* in the word list divided by the frequency of all sounds or by the number of words in the list. By RELATIVE DISTINCTIVENESS is to be understood the “distinctive value” of a phoneme s_1 to another phoneme s_2. The importance of this value is shown on the one hand by the definition of the phoneme concept as the smallest meaning-distinguishing unit, from which it follows that a sound must be distinctive to all other sounds in order to obtain phonemic status. On the other hand, the terms “weak-distinctive” and “strong-distinctive” play an important role in the theory of sound changes. In the theoretical model, weakly-distinctive phonemes may merge with other phonemes without seriously affecting the language as a system. A suitable measure of relative distinctiveness is the number of common minimal pairs in relation to the size of the word list or the number of words containing the sounds s_1 and s_2.

The degree of this influence depends on the type and range of the sound change. In the following theoretical model, the different effects for the types *conditioned-unconditioned* and *merging-shifting* are illustrated. In UNCONDITIONED SOUND CHANGES, the sound change occurs independently of the phonetic environment. This fact leads to an insurmountable obstacle in most methods of IR, such as the morphophonemic and phonotactic methods, so that the distinctive method could fill a missing gap in the field of IR. An unconditioned sound change may lead to a merger of two sounds or a new sound in the phoneme system. As an example, the Proto-Germanic sound **ē₁* is named in this chapter, which in Gothic merged with **ē₂* (cp. Fig. 11.1). In North- and West-Germanic, this sound was lowered to the new phoneme **ā*, which, apart from a nasalized long vowel, had no longer existed in Germanic after the change **ā* > **ō*. Both sound changes led to different changes in the distinctiveness of the affected phonemes. SHIFTING does not lead to any change in distinctiveness — neither in the absolute distinctiveness of the phoneme **ē₁* nor in the relative distinctiveness to other sounds. Of more importance is the MERGING of two sounds. It results in a greater degree of distinctiveness for the newly formed phoneme. The merger also leads to an increase in homophony, as long as the relative distinctiveness between both sounds has been greater than zero. These changes in the language system can be far-reaching so that it can be theoretically assumed that first and foremost weakly-distinctive sound pairs are affected by merging (as in the case of **ē₂*, which occurred mostly in the loanword vocabulary).

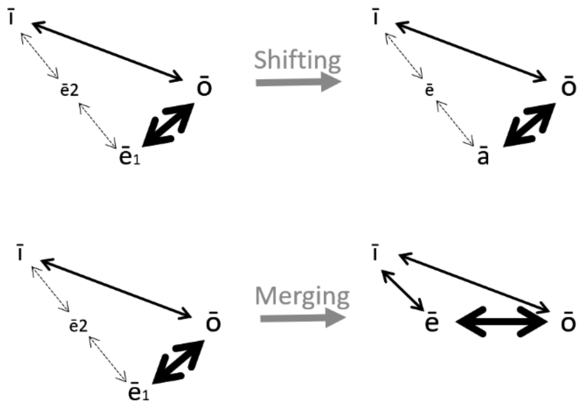

Figure 11.1: On the left, the simplified phoneme system of long vowels in Proto-Germanic (without nasalized long vowels). Absolute distinctiveness is represented by the size of the sound, relative distinctiveness by the thickness of the arrows. The first sample illustrates the North-West-Germanic development as an example for shifting; the second sample shows the Gothic development as an example for merging.

If an unconditioned sound change leads to the emergence of a new sound, then, in its initial phase, this sound stands in an allophonic relation to its originating sound. An allophone can have a comparatively high absolute distinctiveness but at the same time (in the case of complementarity) a relative distinctiveness of zero to the other allophones. Other phonemes have a comparatively low relative distinctiveness to both allophones since it holds that the intersection of two minimal pairs *mp(s, allo-*$phon_1$*)* and *mp(s, allophone*$_2$*)* is equal to zero. If the unconditioned sound change leads to a merger, this results in a higher frequency of the new phoneme and a lower frequency of the old sound.

The allophony that has arisen after a sound change $s_1 > s_2$ / C can be dissolved by new lexemes (e.g., loanwords, neologisms) so it no longer follows this sound distribution. A far greater impact to covering the sound change is also brought about by new sound changes. Here it has to be distinguished whether the sounds s_1 and s_2 or the conditions C are affected by the new sound change. Of particular interest for the distinctive approach are the cases of "neutralized conditions." This applies, among other things, when the conditional sounds disappear or merge with *schwa* or other sounds. Cases of neutralized conditions occur relatively often since information about the phonetic environment was transferred by the first sound change. Well-known examples are *umlauts* and palatalizations, which are often followed by neutralization or loss of the suffix vowels, e.g.:

- OHG *kās**i*** 'cheese' > NHG *Käse* [ˈ**k**ɛːzə]
- Proto-Celtic **wir**ī*** 'man (gen.sg.)' > OIr. *fir* [fi**r**ʲ]

The consequence of these neutralizations is that the sounds s_1 and s_2 occur more frequently in minimal pairs, as in German *m*ü*sste* 'must (1st, 3rd sg. subj. II)' - *m*u*sste* 'must (1st, 3rd sg. prt.)'. This is especially true when working with morphemes instead of wordforms and considering allomorphs separately (e.g., German *schl*ag - *schl*äg 'hit').

11.2 Procedure of the Distinctive Approach

11.2.1 Preprocessing

For machine processing, a (transcribed) list of the wordforms or morphemes of the language in question is required. For instance, such a list can be extracted from the lexicon of the CELEX project of the Max Planck Institute Nijmegen (see Sect.

10.3.3). Among other things, the project offers a German lexicon in which the lexemes are already decomposed into morphemes. Some allomorphs (e.g., German *schläg-(t)* '(he) hits' and *schlag-(en)* '(they) hit') are not abstracted to a common morpheme, which can be important for further usage. However, not each lexeme is incompletely decomposed (e.g., *verschlag-en*), giving the final morpheme list a broader scope than morpheme lists from other sources.

Finally, the obtained morpheme list is transcribed into IPA or XSAMPA using a converter. Transcription errors can occur during this process since converters are usually implemented for the conversion of words. In German, for example, the *schwa* sound of suffixes may not be recognized as such, since the morpheme considered in isolation is not recognized as a suffix by the converter. The glottal stop, which is often missing at morpheme boundaries, has been omitted entirely.

For each phone in the morpheme list, a separate set of morphemes containing the corresponding sound is created (called *phone-morpheme set* in the following), where the respective sound is replaced by a placeholder. If a sound occurs twice in the same morpheme (e.g., *a* in *Lama*), the morpheme appears twice in the phone-morpheme set: Once as *L_ma* and once as *Lam_a*. The morpheme *L_ma* then appears both with the phone *a* (*Lama*) and with *i* (*Lima*), e.g.,

- *W(a) = { L_ma, Lam_a, _lt }*
- *W(i) = { L_ma, _nsel }*

The size of the set represents the absolute frequency of the sound (*freq(a)*). The placeholder makes it easier to find common minimal pairs because the intersection of two phone-morpheme sets gives the minimal pairs of these phonemes (*mp(a,i)*).

$$freq\ (a) = |W(a)| = |\{\ L\_ma, Lam\_, \_lt\}| = 3$$

$$mp\ (a, i) = W(a) \cap W(i) = \{L\_ma\}$$

11.2.2 Determination of Complementary Sounds

For two allophones *i* and *j*, the theory holds: $M_i \cap M_j = 0$ with M_i is the phone-morpheme set of the allophone *i*. This results from the fact that complementarily distributed allophones occur in a different environment and thus no minimal pairs can be formed. The minimal pairs also provide information about the sound classes of the phones. Since the sounds in a word are not randomly distributed, different intersections are to be expected. For example, vowels occur more often in *inlaut*-positions and consonants at the syllable edges. This results in a larger intersection of vowels with other vowels than with consonants and in a larger intersection of liquids with each other. In theory, therefore, the following relations apply (with A as a sound class, *i,j* as phones and /i/ as a phoneme):

$$\left(M_{i\,\in\,A} \cap M_{j\,\in\,A}\right) > \left(M_{i\,\in\,A} \cap M_{j\,\notin\,A}\right) \geq \left(M_{i\,\in\,/i/} \cap M_{j\,\in\,/i/}\right) = 0$$

It should be noted, however, that the absolute values are rather unsuitable for this purpose since a smaller number of minimal pairs is naturally to be expected for rare phones, even if they are of the same sound class. For the normalization of intersections, the Jaccard index (see Sect. 8.1.2.4) is suitable, which is widely used in data mining (cf. Liu 2007:138). It divides the size of the intersection set by the size of the union set and thus also establishes the similarity of two sets. However, since the size of the intersection is often zero even for sounds from other sound classes or for rare sounds, information about the size of the two sets gets lost in this way. Addition of the dividend with one prevents a division of zero (so-called Add-One-Smoothing):

$$\frac{\left|\left(M_{i\,\in\,A} \cap M_{j\,\in\,A}\right)\right| + 1}{\left|\left(M_{i\,\in\,A} \cup M_{j\,\in\,A}\right)\right|} > \frac{\left|\left(M_{i\,\in\,A} \cap M_{j\,\notin\,A}\right)\right| + 1}{\left|\left(M_{i\,\in\,A} \cup M_{j\,\notin\,A}\right)\right|} \geq \frac{\left|\left(M_{i\,\in\,/i/} \cap M_{j\,\in\,/i/}\right)\right| + 1}{\left|\left(M_{i\,\in\,/i/} \cup M_{j\,\in\,/i/}\right)\right|}$$

Since allophones usually belong to the same sound class (i.e., if *i,j* ∈ /i/ and *i* ∈ A, then also *j* ∈ A), we can determine for sound *i* that the sound *j* is its allophone if for *j* holds:

$$R(i,j) := \min_{j\,\in\,A} \frac{\left|\left(M_{i\,\in\,A} \cap M_{j\,\in\,A}\right)\right| + 1}{\left|\left(M_{i\,\in\,A} \cup M_{j\,\in\,A}\right)\right|}$$

Tab. 11.A illustrates an over-dominance of low-frequent sounds (e.g., ([dʒ, tʃ, j] or [ŋ]). This arises from the high size of the set of the second, high-frequent sound, which causes the denominator to reduce the result of the division.

Table 11.A. The lowest values of R(i,j) for the German complementary pairs *x-ç* and *s-z* (with data from the CELEX corpus).

[x]		[ç]		[s]		[z]	
x-b	0.00134	ç-g	0.00053	s-t̲ʃ̲	0.00059	z-tʃ	0.00103
x-h	0.00144	ç-v	0.00076	s-d̲ʒ̲	0.00060	z-dʒ	0.00104
x-ç	*0.00147*	ç-h	0.00104	s-h	0.00093	*z-s*	*0.00114*
x-z	0.00172	ç-j	0.00122	*s-z*	*0.00114*	z-ŋ	0.00155
x-j	0.00182	*ç-x*	*0.00147*	s-ʒ	0.00118	z-x	0.00172

11.2.3 Clustering of Complementary Sounds

Similar to the phonotactic method (see Sect. 10.2.4.1), sounds can be clustered based on their similarities to each other to illustrate sound groups that have more frequent minimal pairs than others and thus have greater distinctiveness to each other. Since true allophones do not have any minimal pairs with each other, these can become partially visible. *Clustering* refers to different techniques for determining sub-groups (so-called *clusters*) within data sets (in our example, the sounds). In contrast to *classification*, where the sub-groups (e.g., vowels and consonants) are already known, the number of clusters is unknown and have to be identified. Each identified cluster contains data that can be formally distinguished from data of other clusters. Since our data are based on minimal pairs, vowels have more minimal pairs to each other than consonants for phonotactic reasons. As expected, consonants and vowels form their own clusters in this way.

A common unsupervised machine learning clustering algorithm is *k-means clustering* (Lloyd 1982). This method generates *k* clusters, each of which has a center point (i.e., *centroid*) corresponding to the average of the points in the cluster. Before the procedure, the data must be standardized to make the variable comparable. This involves adjusting the values so that the mean is zero and the standard variation is one. The standard procedure was developed by Hartigan and Wong (1979) and defines the variation within a cluster as the sum of squared Euclidean distances between the cluster's elements and the centroid:

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad \begin{matrix} \textit{with } x_i \textit{ as point of the cluster } C_k \\ \textit{and } \mu_k \textit{ as centroid of } C_k \end{matrix}$$

The clusters are to be defined in such a way that the variation within a cluster is as minimal as possible. For a detailed description of the procedure or other cluster algorithms, I refer here to MacKay (2003:285–290) and Weiss et al. (2010).

The *k*-means algorithm assumes that the number of clusters is already known. Since this is normally not known in advance, there are different methods to determine the optimal number of clusters. The ELBOW METHOD computes the so-called *total within-cluster sum of square* (*wcss*) for all possible values of *k* (i.e., the sum of $W(C_k)$ of all clusters). A diagram with all *wcss*-values for each cluster number *k* shows the best cluster number that is defined as the sharp point looking like an elbow. The SILHOUETTE METHOD determines how well each data point lies within its cluster (Rosseeuw 1987). The average value for all (i.e., *average silhouette*) determines the quality of cluster analysis. The best cluster number is defined as the value that maximizes this average silhouette for *k*. The third method is the GAP STATISTICS

METHOD, which compares the total intra-cluster variation for each *k* with the expected values under the null reference distribution of the points using Monte Carlo simulations (for details, see Tibshirani et al. 2001). For the German test data, the methods indicate a k-value of two to three as “optimal.”[72]

11.2.4 Interpretation of Clusters

App. E.1 shows a network of German sounds with undirected graphs weighted by the Jaccard index. A high Jaccard index between two sounds leads to spatial proximity in the network. Better clustering results can be achieved by rejecting pairs with a low Jaccard index, which also leads to the isolation of low-frequent sounds (e.g., [ʒ, dʒ]). In Fig. 11.2, a threshold was set for this purpose, which was one-fourth of the average Jaccard index. Jaccard indices under this value were set to 0.0. The network is characterized by the phonetic structure of a morpheme, which often resembles the syllable structure *CVC*. This leads to clusters with sounds frequently occurring at the same position of syllables.

The vowels form four clusters (Fig. 11.2, second picture), of which the nasalized vowel [ã] forms a separate cluster due to the lack of minimal pairs. The sounds [ə] and [ɐ] only occur in unstressed syllables, which is the reason why they hardly form minimal pairs with other vowels. The short vowels form the first cluster and the long vowels and diphthongs the second cluster. Inferences about diachronic sound change can be drawn from the way the sounds form clusters. The separation of short and long vowels results from the Early New High German lengthening of short vowels in open tone syllables (Paul et al. 1969:52 §23). This has resulted in short vowels only occurring in closed syllables. The clustering of centralized vowels [ə, ɐ] are due to the Old High German weakening of unstressed syllables (Paul et al. 1969:61 §27). In this case, all the vowels have merged into [ə]. Despite these cases, inferring SOUND CHANGE FROM CLUSTERING is difficult. Thus, the clusters of consonants cannot be directly attributed to sound change.

The two large consonantal clusters reflect the phonotactic structure *CVC* (Fig. 11.2, first picture). The first cluster (yellow) shows those consonants that are not usually found at the end of syllables in German (e.g., [h, j, v]), the second (red) those that are rarely found at the beginning of sounds (e.g., [ç, x]). The largest number of German consonants occur frequently at both syllable edges so that they are posited

[72] The optimal number is between two (silhouette method) and three (elbow and gap statistics method) clusters.

at the intersection area. The network reveals further phonetic groups inside the clusters. This includes the voiced plosives (App. E.2), which are posited at the edge of the intersection area since this is where final devoicing begins to take effect. Likewise, the liquids occupy a special area, since they are more frequently posited between consonants and vowels.

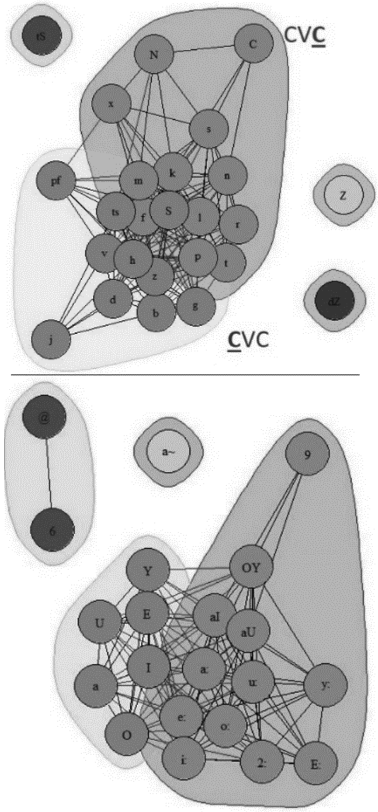

Figure 11.2: Result of the cluster algorithm with the German data of the CELEX project. Ties with below-average Jaccard index were removed. The sounds were transcribed into XSAMPA. The first illustration shows the result for all sounds, the second for consonants and the third for vowels.

Consequently, the network primarily reflects the phonotactic structure of German morphemes. This information is important for identifying allophones and sound pairs since they can be posited at different positions in the network depending on the distribution's rule. To illustrate this, let us sketch the development of the Germanic sound **x* in a highly simplified way: Germanic **x* developed from

Proto-Indo-European *k and *ḱ and occurred both at the beginning and end of morphemes, so it was therefore posited in the intersection of the two consonantal clusters. Through Verner's law, the sound [ɣ] split off, which then merged with the old Proto-Indo-European *g^h. This had the effect of moving *x closer to the **C**V-cluster. The Old High German lenition caused *x to become *h* at the initial position and between vowels. As a result, both allophones left the middle of the intersection area. The merging of *x* with the velar fricative from the High German consonant shift, as well as the loss of [h] in the *inlaut* position, brought the two old allophones to the edges of the clusters.

11.2.5 Inferring Sound Changes from Clusters

Since the resulting clusters represent typical structures of a language, potential sound changes have to be inferred differently, depending on whether the condition of the sound change conforms to the clusters' structure (*isomorphic*) or is opposite to it (*heteromorphic*). An isomorphic condition is present for the consonantal *C*V*C* structure if the sound change leads to an imbalance of *C*V and V*C*. The result is that the two complementary sounds move to the edges of the corresponding cluster. Examples are the Vulgar-Latin palatalization of *k* before front vowels or final devoicing in Russian.

In all other cases, there is a heteromorphic condition that causes both sounds to remain in the same cluster but positioned away from each other. Examples are the German allophones of /x/ (*ç* > *x* / _[+back vowel]) and unconditioned sound changes.

11.2.5.1 Complementary Sounds Within the Same Cluster (Heteromorphic Conditions)

Table 11.B. Result The top-ranked complementary sounds of pairs within the same clusters (performed with German data from the CELEX corpus). For the calculation of the cosine similarity, the sound vectors with Jaccard indices are used. The allophones of /x/ are highlighted.

Rank	Pair	Cosine similarity	Jaccard index
1	d-p͡f	0.86161	0
2	*ç-x*	*0.85100*	*0*
3	œ-yː	0.80873	0
4	t͡ʃ-z	0.78819	0
5	j-p͡f	0.76641	0

If the conditions of a sound change are heteromorphic, both complementary sounds will remain in the same cluster and move away within the cluster. This is true, for example, of the development of the palatal allophone [ç] from /x/. The relations of *ç* and *x* to the other sounds remained largely unchanged. This leads to the paradox that the cosine similarity of allophones within a cluster is very high, while the Jaccard index is zero. Low Jaccard indices are much rarer within a cluster than between sounds of different clusters and are therefore particularly useful for identifying potential sound pairs. Tab. 11.B lists the highest cosine similarity values with a Jaccard index of zero. Despite the high similarity of the allophones *ç* and *x*, this one shows only the second-best value, which is due to the fact that especially low-frequent sounds such as *p͡f*, *j* or *t͜ʃ* may not have minimal pairs with another sound, even purely by chance. Since, as can be seen from Fig. 11.2, the allophones *ç* and *x* have a stronger distance within the cluster, false positives may overtake true allophone pairs in the ranking.

It is therefore advisable to include the potential conditions in the process. An approach with sound classes is presented here as an example. For this purpose, the co-occurrence of a sound s_1 and a sound class y is calculated:

$$conDiff(s_1, s_2) := \frac{\sum_i freq_{s_1 \cup s_i}}{freq_{s_1}} - \frac{\sum_i freq_{s_2 \cup s_i}}{freq_{s_2}} \; with\ i \in sound\ class\ y$$

By considering the conditions, the top-ranked pair *d-p͡f* moves to the back. However, it should be noted that due to the necessary condition of a Jaccard index of zero, only allophones can be detected with this approach. This circumstance is the reason why the distinctive method is not of interest for IR of sound change with heteromorphic conditions. The focus of the method is therefore on the reconstruction of sound changes with isomorphic conditions.

Table 11.C. The top-ranked sound pairs with conditions (performed with the same data from Tab. 11.B). The final measure weights its factors (evaluation of conditions and cosine similarity) in favor of the cosine similarity ($g_1 = 1.0$, $g_2 = 5.0$). The following sound classes were considered: vowel, consonant, long vowel, short vowel, front vowel, diphthong, back vowel, high vowel, mid-vowel, open vowel, schwa, voiced consonant, voiceless consonant, plosives, affricates, fricatives, nasals, liquids, glides, labials, dental, palatals, and velars. True positives are highlighted. The correct condition of *t͡ʃ-d͡ʒ* would be _#.

	conDiff(s_1,s_2)	cosSim(s_1,s_2)	conDiff × g_1+ cosSim × g_2
ç - x with front vowels	*1.23724*	*0.85100*	*5.49223*
x - ç with back vowels	*0.93294*	*0.85100*	*5.18793*
d-pf with vowels	0.83680	0.86166	5.14488
j - pf with dentals	1.16205	0.76641	4.99408
t͡ʃ-d͡ʒ with consonants	*1.06667*	*0.73973*	*4.76532*

11.2.5.2 Complementary Sounds from Different Clusters (Isomorphic Conditions)

If the distribution of complementary sounds conforms to the syllable structure of the clusters, they may be in different clusters. This fact leads to a lower cosine similarity of complementary sounds and a different methodological approach. In order to demonstrate the behavior of sounds after a sound change, two simulated sound changes were performed on the data set. The first simulated sound change replaced each *l* before a front vowel with a palatal *ʎ*. The sound change affects the distribution of laterals within the syllable structure. The palatal sound only appears in the prevocalic position in morphemes of the *CV* structure, whereas the alveolar sound occurs less frequently in this position and thus is overrepresented in the postvocalic position. Fig 11.7 illustrates the splitting of the phoneme. While the sound *l* originally appeared within the intersection area of both consonantal clusters (see Fig. 11.2), a horizontal split occurs towards the cluster edges after the sound change (Fig. 11.3).

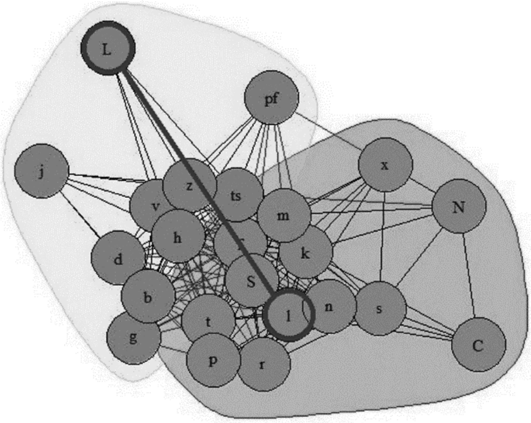

Figure 11.3: Cluster network of the data from Fig. 11.2 after applying the simulated sound change *l* > *λ* / _V[+front]. Both lateral sounds move towards the edges of their clusters.

The second sound change simulated a palatalization of [k] before and after high front vowels (k > c / V[+front, +high], see Fig. 11.4). The syllable structure is hardly affected by this condition so that the central position of *k* is not changed. The newly formed palatal is a low-frequent sound and, just like them form its own cluster.

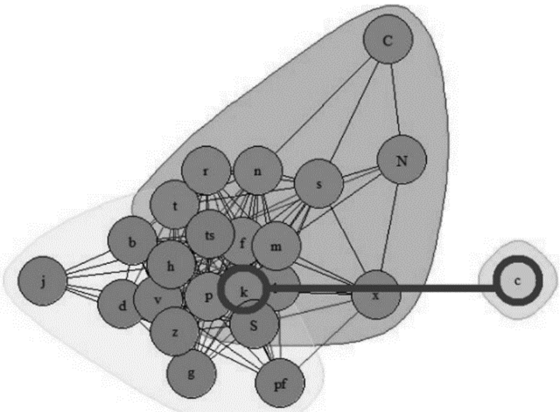

Figure 11.4: Cluster network of the data from Fig. 11.2 after applying the simulated sound change *k* > *c* / V[+front, +high]. The vector of *k* remains in the intersection area, while [c] was pushed to an own cluster.

A Jaccard index of zero gives an indication of a potential sound pair due to its rarity within a cluster. However, low Jaccard indices tend to be the rule for sounds of different clusters. This leads to the case that several potential complementary sounds are available for one sound. Therefore, the correct identification can only be solved by including further information. For a linguist, there may be certainly a preference for *λ* without the knowledge about the simulated sound change since this sound is the only lateral in the list of potential complementary sounds. Consequently, three factors serve to determine sound pairs from different clusters:

- both sounds belong to the same sound class but different clusters
- they have a low Jaccard index
- they should be phonotactically similar

As with the determination of sound pairs within the same cluster, the potential condition should also be taken into account. Since these depend less on the phonetic and more on the syllable-structural environment, "structural conditions" are defined here. In the concrete example, these are the structures *initial sound*, *final sound*, *pre-consonantal initial sound* and *post-consonantal final sound*. A Jaccard index below a threshold value of 0.01 prevents sound pairs of a vowel and consonant. A low Jaccard index also plays a role as a potential summand. Cosine similarity is suitable to implement the third condition. The following formula $JaccCos(s_1,s_2)$ was used in this work (with *g* as weights). An alternative to cosine similarity is to include empirical probabilities for a sound correspondence as summand in the formula (i.e., $JaccTypo(s_1,s_2)$). For this purpose, the data of Brown et al. (2013) as presented in Sect. 10.2.4.4 was used.

$$JaccCos(s_1, s_2) := g_1 \times conDiff(s_1, s_2) + g_2 \times cosSim(s_1, s_2)$$

$$JaccTypo(s_1, s_2) := g_1 \times conDiff(s_1, s_2) + g_2 \times typoProb(s_1, s_2)$$

By restricting the potential complementary sounds to sounds of the same sound class, the number of potential sound pairs may be greatly reduced. As can be seen from App. E.3, single, particularly centrally posited sounds dominate the result list, which makes an additional selection procedure necessary. The following ITERATIVE ALGORITHM allows only one complementary sound for each sound, so that, conversely, the procedure cannot identify two sound changes with the same sound:

1. Select the sound pair s_1,s_2 with the highest measure JaccCos(s_1,s_2)
2. Remove all sound pairs containing the sounds s_1 and s_2

3. Repeat the process until a set endpoint (threshold or number of candidates) is reached

The most relevant sound pair with the highest *JaccCos* value in App. E.3 (i.e., *ʃ-ç in* #_) is selected (item 1) and all other sound pairs containing *ʃ* or *ç* are deleted (item 2). This process is repeated until each sound of the smaller clusters has a complementary sound or the threshold is reached.

Table 11.D. Result of the algorithm of Sect. 11.2.5.2 using German data from the CELEX corpus, the clusters of Fig. 11.2, and the JaccCos formula. On the right, the corresponding historical sound changes. The complementary sounds for which the sound change direction was correctly identified are highlighted in italics. The sound changes with wrong conditions are marked with (≠).

	Sound change	**JaccCos**	**Historical correspondence**
consonants	ç > ʃ / #_	1.25600	
	d > t / _#	*1.24358*	final devoicing: d > t / _#
	p > b / #_ (≠)	*0.97822*	final devoicing: b > p / _# (mis-interpreted as *anlaut* lenition)
	s > z / #_	*0.92623*	OHG spirant lenition: s̠ > z̠̊ / V_(R)V)
	x > v / #_	0.89074	
	g > k / _#	*0.88790*	final devoicing: g > k / _#
vowels	iː > ʏ / #C_	1.20891	
	ɛ > øː / _C#	1.04817	
	*oː > ɔ / #C_* (≠)	*1.03873*	lengthening in open tone syllables: o > oː / _${-t, -m, -er}
	a > ɐ / _#	1.01452	
	œ > ɪ / _C#	0.96160	
	uː > ʊ / #_ (≠)	*0.95371*	lengthening in open tone syllables: u > uː / _${-t, -m, -er}
	eː > ə / _# (≠)	*0.84237*	weakening of unstressed vowels: e > ə / _[-stressed]

The result varies depending on which clusters the *k*-Means algorithm arrives. The method using the clusters of Fig. 11.2 and German CELEX data arrives at the sound changes in Tab. 11.D. Of these pairs, seven out of thirteen can be assigned as sound change, giving a MATCH RATE of 0.5385. The precision of the phonological rules is less than half (see Tab. 11.E). If we restrict ourselves to the phonological rules with "structural condition," a recall of 0.3846 can be achieved. For consonants, the method correctly detects three out of six recognized sound changes. The final devoicing of labial plosives is falsely recognized as *anlaut* lenition. In the case of vowels, the procedure identifies the lengthening of short vowels in open tone syllables, but the sound change direction could not be determined correctly. From seven diachronic sound changes, four belong to the Middle-High-German period, two to the New-High-German period and one to the stage of Old High German.

To illustrate further sound changes, the procedure was repeated with the data set which contains the two simulated sound changes *l* > *ʎ* before front vowels and *k* > *c* before and after high front vowels. The lateral sound change shows a migration of the affected sounds to the edges of the clusters and ranks third (see Fig. 11.4). The palatalization of *k* is identified at rank seven. Since this sound change contains *k*, the procedure can no longer identify the final devoicing of *g* > *k*.

Table 11.E. Precision, recall and F-score for the match rate of the distinctive method with German data. Phonological rules and sound changes are listed separately.

Match rate	Phonological Rules	Sound Changes
Precision	0.23077	0.53846
Recall	0.11539	0.03241
F-score	0.15385	0.06114

The tested means for identifying the direction (i.e., the dominance of the new sound) are not suitable as such. The change eː > ə / _# reflects the sound change V > ə in unstressed syllables, which in German mostly applies to the last syllable and thus frequently in final position. Overall, the method was able to correctly determine the sound change direction in four of the seven matched pairs (RATE OF CORRECT CHANGE DIRECTION = 0.5714). Since the method allows only one sound pairing for each sound, the number of final sound pairs is significantly smaller than in other methods. This reduces the number of false positives and a relatively high precision value of 0.5385 can be achieved (see Tab. 11.E). At the same time, this also significantly reduces the recall.

The condition was largely provisionally treated in this method. Only positions that conform to the clusters' structure were possible (i.e., #_, _#, #C_, and _C#).

With the exception of the lengthened vowels, a correct phonetic environment was determined in five of the seven correctly identified sound changes (RATE OF CONDITIONS = 0.7142). This good rate results from the choice of conditions laid. The choice of other conditions can consequently lead to other sound changes.

Table 11.F. Result of the distinctive method for German consonants with different phonetic classes (CELEX corpus, JaccCos, $g_1 = 1.9$, $g_2 = 1.4$). The complementary sounds with correct sound change direction are in italics. The sound changes with wrong conditions are marked with (≠).

Manner of articulation	**Manner and place of articulation**	**Manner and place of articulation and voice**	**Dolgopolsky classes**
d > t / _# (1.57249)	*d > t / _# (1.57249)*	t > h / #_ (1.68491)	*d > t / _# (1.57249)*
ç > ʃ / #_ (1.53194)	ç > ʃ / #_ (1.53194)	ç > ʃ / #_ (1.53194)	ŋ > h / #_ (1.33686)
p > b / #_ (≠, 1.29969)	f > pf / _# (1.40425)	r > b / #_ (1.45657)	*p > b / #_ (≠, 1.29969)*
g > k / _# (1.18552)	n > ts / #_ (1.39832)	p > ts / _# (1.45423)	*g > k / _# (1.18552)*
s > z / #_ (1.15572)	r > z / #_ (1.38226)	d > l / _# (1.42939)	*s > z / #_ (1.15572)*
x > v / #_ (1.09253)	*p > b / #_ (≠, 1.29969)*	f > pf / _# (1.40425)	
	g > k / _# (1.18552)	n > z / #_ (1.37666)	
	s > v / #_ (1.14583)	j > m / _# (1.27260)	
		g > k / _# (1.18552)	
		s > v / #_ (1.14583)	

An open point is the definition of the sound classes. Due to the restriction of sound pairs to the same articulation type, only sound changes with this restriction are possible. Since a large number of the non-assimilatory sound changes occur at this level, they can be identified in this way, but others are excluded in advance. Tab. 11.F shows the results with several phonetic classes. As an alternative, Dolgopolsky's (1964) classes provide good results, but a class usually consists of a few sounds.

To compare the two formulas *JaccCos* and *JaccTypo*, the procedure was performed with the German IPA data from Wiktionary and was repeated with ten different clusters. Using the *JaccTypo* formula achieves better precision, but the difference is not significant (*JaccTypo* with 0.2174 and *JaccCos* with 0.1910).

11.3 Discussion

The test was able to identify some sound changes in German and shows the basic possibility of the method. These include the allophone pair *ç* and *x*, the final devoicing, and the simulated examples. A strength of this approach is the easier identification of complementary sounds compared to the phonotactic method by using clustering. From the absence of a final *g* and a high frequency of final *k*, a sound change cannot be concluded immediately, since other sounds also behave phonotactically similarly (cf. Sect. 4.2.2). The correct sounds tend to form fewer minimal pairs, since there are no or only a few minimal pairs in relevant environments. However, with minimal pairs there is significantly less distortion in the irrelevant environments than with phonotagms. For example, the allophones of German /x/ do not form any intersection with each other, while phonotagms such as _ə occur for both sounds. The distinctive method is therefore a special variant of the phonotactic method which works on the word level.

A disadvantage of the method is the low time depth of the identified sound changes. These are largely phonological rules of the synchronic language (e.g., the change *s* > *z* #_). In addition, there is a restriction of possible sound changes, caused on the one hand by single conditions (e.g., the same sound class) or the absence of a zero morpheme. Similar to the phonotactic method, this approach also has difficulty in predicting the correct sound change direction. The correct interpretation of the clusters poses another difficulty. Although diachronic changes between the clusters could be traced with the simulated sound changes, questions remain for further research, including the influence of different conditions of sound changes and the question of the "original" cluster structures (i.e., when do sound changes create new clusters?).

11.4 Application Case: Proto-Indo-European

To perform the distinctive method on Proto-Indo-European data, the resources from Sect. 9.4.1 were used, consisting exclusively of verb roots. Since only the vowels *e* and *a* appear in the data set, the vowels are ignored in the clustering procedure.

11.4.1 Clustering

The determination of the optimal consonant clusters using all three methods (elbow method, silhouette method, and gap statistics method) proposes the value 2. The result shows a cluster with the sound [ḱ^h, p^h, t^h] and another cluster containing the remaining consonants. The aspirated *tenues* occur (almost) exclusively after initial *s* and are to be understood as allophones of the aspirated *mediae* (so-called Siebs's law, cf. Rasmussen 1989:154). The clusters, therefore, reflect a phonological rule of the proto-language.

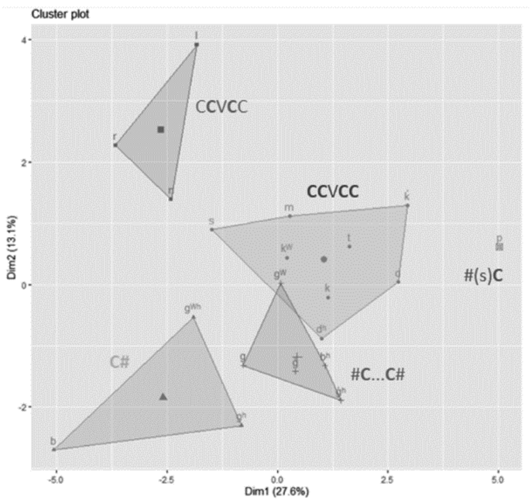

Figure 11.5: Result of a k-means-clustering with Proto-Indo-European data after the removal of vowels and *tenues aspiratae*: The five created clusters form phonotactical clusters.

To gain deeper insight, clustering was repeated with the data without the aspirated *tenues*. The elbow method suggests five optimal clusters, the silhouette method two, and the gap statistics method only one cluster. The five optimal clusters are shown in Fig. 11.5 and can be interpreted phonotactically. The cluster (CCVCC) reflects sonorants (*r, l*, and *n*) and occur frequently in positions before and after vowels. The sounds of the cluster C# (*b*, *g*^{h}, and *g*$^{u̯h}$) are rarely found in initial positions and before vowels. The cluster #C…C# (*$g^{u̯}$, g, $ǵ^h$, ǵ*, and *b^h*) includes sounds that are frequently found in initial or root-final position but less frequently in middle positions. The central cluster includes sounds that are more common to each of these positions. The special role of *p* results from its dominant positioning at the beginning of a syllable or in *#sp-*. In the case of only two clusters, the sounds are grouped according to root-initial and root-final consonants.

11.4.2 Determination of Complementary Sounds

The first clustering already indicates a sound change that leads to the emergence of the *tenues aspiratae*. For the determination of the complementary sounds, all sound classes with the same articulation manner and place were considered. The pairs with the highest *JaccCos*-values are *p^h-b^h* / #C_, *$ḱ^h$-$ǵ^h$* / #C_, and *t^h-b* / #_ (cf. Tab. 11.G). The correct sound pairs are *p^h-b^h, $ḱ^h$-$ǵ^h$*, and *t^h-d^h* with the condition #s_. The best recall is achieved with a weighting of $g_1 >$ 0.8 and $g_2 <$ 1.4 (i.e., a high weighting for *conDiff* and a low weighting for cosine similarity). The correct condition would only be possible with a detailed consideration of the phonetic environment, which is more likely to be achieved by the phonotactic approach. The incorrect determination of the pair *t^h-b* results from the divergent use of *t^h* that also occurs in the initial position in the data set and is thus more likely to be complementary to *b*, which does not occur in that position.

Table 11.G. Result of algorithm for determining complementary sounds using Proto-Indo-European data: in the upper part, all consonants are considered and in the lower part, the aspirated tenues are removed. All sound classes with the same articulation manner and place were considered. The weights g_1 and g_2 were set to 2.0 for conDiff and to 0.5 for cosSim.

			Pair	**conDiff**	**cosSim**	**JaccCos**
with	aspirated *tenues*	1	p^h-b^h in #C_	1.0	0.44366	2.22183
		2	$ḱ^h$-$ǵ^h$ in #C_	1.0	0.25584	2.12792
		3	t^h-b in #_	0.5	0.21439	1.10719
without		1	g^W-b in #_	0.575	0.63201	1.20701
		2	g^h-k^W in _#	0.30702	0.80338	1.11039

Without *tenues aspiratae*, the procedure identifies with both two and five clusters the sound pairs *$g^{u̯}$-b* at the initial position and *g^h-$k^{u̯}$* at the final position. The algorithm attempts to look for pairs in opposing patterns, and thus for sound changes that conform to the clusters' structure. The first sound pair $g^{u̯}$-*b* is relevant due to the gap of *b* in the initial position. The sound $g^{u̯}$ is chosen as complementary sound because it was less frequent in the final position compared to other sounds in the same cluster. The second pairing was formed in a similar way (g^h as initial sound and $k^{u̯}$ as root-final sound). However, whether these sound correspondences are causally related may be doubted. Rather, there are alterations between *tenues* and *mediae aspiratae* in Proto-Indo-European (cf. Meier-Brügger 2010:269), so that both sound pairs can be considered false positives. The reason for the gap of *b* cannot be determined internally. The rarer occurrence of labiovelars after vowels can also be explained by the βουκόλος rule and the delabialization before *t* (e.g., **lei̯*$k^{u̯}$- > **lik-to*). Both alternations are context-dependent and do not depend on the position of the root structure. Proto-Indo-European lacks examples of such position-dependent alternations that could be identified by the algorithm,[73] so no recall can be determined here.

11.4.3 Discussion

The approach is able to identify the synchronic Siebs's law, as well as to give evidence for the *b*-gap. As in the case of German, these are synchronically recognizable rules or distributions, so that the issue of low time-depth is also true for Proto-Indo-European. For proto-languages, there is the additional fact that the method can only reconstruct what the linguist implements in his reconstruction. In LIV, Siebs's law is considered despite its allophonic nature, while voiced [z] as an allophone of /s/ does not appear in the entries of the Proto-Indo-European dictionary. Such differences make it clear that IR applied to proto-languages, in many ways, reflects the linguist's picture of the proto-language's history and does not generate a new picture.

[73] An example of such position-depended alternations is the Proto-Indo-European final (de)voicing (Fortson 2010:80), which is not implemented in LIV.

12. Gap Method

The gap approach is an internal reconstruction method that is proposed in the literature to reconstruct unconditional sound change internally (cf. Sect. 4.2.2.3). For the reconstruction of unconditional sound changes, sound environments do not play a role but the distribution of sounds in the phoneme system. The gap method considers gaps or asymmetries in the system and assumes a sound change as their cause.

In order to implement the approach computationally, the sounds of a phoneme system are broken down into their individual distinctive features and an “ideal phoneme system” is formed on the basis of these features. In the ideal phoneme system, all possible combinations are considered for all distinctive features. If a language has a bilabial plosive, a bilabial nasal, and a velar plosive, the ideal phoneme system also requires a velar nasal. If this sound is missing in the actual phoneme system, the gap is attributed to a theoretical sound change. How valid this assumption holds, is a question to be answered in the evaluation section.

The method is not applied to Proto-Indo-European data because it does not provide optimal conditions for proto-languages. The gap approach requires precise knowledge of the pronunciation of the sounds, which is especially not the case for the laryngeals. They are often interpreted as (guttural) fricatives (cf. Tichy 2009:32), as [h,x, ɣw] (Rasmussen 1994), or even as glottal sounds [ʔ,ʕ,ʕw] (Beekes 1989), in which symmetrical considerations regarding the plosive guttural series may more or less be included. Symmetrical considerations of this kind often play a role in proto-language reconstruction which is another reason why the gap method is rather to be discouraged for proto-languages.

12.1 Procedure of the Gap Approach

12.1.1 Preprocessing

Table 12.A. Sample of determining an opposition (*velar:alveolar*) from sounds that only differ in one feature ([n] and [ŋ]).

[n]	voiced	*velar*	nasal	consonant
[ŋ]	voiced	*alveolar*	nasal	consonant
↓ **velar : alveolar**				

As input data, the approach requires a list of all sounds that the language in question contains, specifying all distinctive features of each sound. To be able to form a phoneme system as ideal as possible, all phones must be described with both the positive and the negative distinctive features. The phone [a] is, for instance, not described as "open front unrounded vowel" but as "unrounded, open, front, *oral*, *short*, vowel" in order to include the oppositions *oral:nasalized* or *long:short* in the phoneme system. All sounds that only differ in one feature are compared. This comparison determines oppositional features, as can be seen in Tab. 12.A. From these opposition pairs, *opposition chains* can, in turn, be formed which correspond to the articulation series (e.g., voicedness, articulation place or manner). For instance, the opposition chain *bilabial - labiodental - alveolar - palatal - velar - glottal* results from the following pairs:

- velar : alveolar
- velar : glottal
- bilabial : alveolar
- labiodental : velar
- velar : palatal

The resulting chains serve as step for the generation of an "ideal" phoneme system. Each feature of each sound is replaced by a feature from the same opposition chain and the result is considered as a possible sound. This results in an ideal system that can be compared with the actual system (as in Tab. 12.B).

Table 12.B. Illustration of an ideal phoneme system for German voiced consonants. Sounds that exist in German are marked with X.

[voiced, consonant]	bilabial	labio-dental	alveolar	post-alveolar	velar	glottal
fricative		X	X	X	X	X
plosive	X		X		X	X
affricate	X		X	X		
nasal	X		X		X	
lateral			X			
trill			X			

12.1.2 Determination of Complementary Sounds

The procedure tries to derive sound change from the difference between the ideal and the actual phoneme system. For this, it is necessary to determine which rows or columns can be combined so that the number of gaps can be reduced. In Tab. 12.B, for example, the column “bilabial” can be combined exclusively with the column “labiodental,” since all other columns already occupy one of the three positions *plosive*, *affricate* or *nasal*. The algorithm works with a gap series like { Ø, fricative, voiced, consonant } for the first row in Tab. 12.B or { bilabial, Ø, voiced, consonant} for the first column. These gap series are compared provided two conditions. First, when the series belong to the same opposition chain. For instance, this is true for { Ø, fricative, voiced, consonant } and { Ø, plosive, voiced, consonant } but not { Ø, fricative, voiced, consonant } and { velar, Ø, voiced, consonant }. Second, both series should not have already occupied the same positions. The gap series { Ø, fricative, voiced, consonant } and { Ø, plosive, voiced, consonant } in Tab. 12.B have three positions the same (alveolar, velar, and glottal) and are thus ruled out for sound change, while { bilabial, Ø, voiced, consonant} and { labiodental, Ø, voiced, consonant} each occupy different positions. The gap series compared in this way are considered as potential sound changes. That one with fewer gaps is identified as the pre-sound of the sound change. Thus, the direction of the sound pair { bilabial, Ø, voiced, consonant } and { labiodental, Ø, voiced, consonant } is *bilabial* > *labiodental*. If one position of the “pre-gap series” is unoccupied and occupied in the other series, this sound is determined as the original sound:

- [**bilabial**, fricative, voiced, consonant] > [**labiodental**, fricative, voiced, consonant]
- [ɸ] > [f]

The procedure is carried out iteratively, that is, after each pass, a sound change is determined, and its gap series is removed for the next iteration step. After each iteration, only the resulting sound (in the sample [f]) is removed, not the pre-sound, since a wrong pre-sound could lead to further wrong pre-sounds in the next iterations. For the evaluation of the sound change, the following formula is applied, which is based on the number of gaps in the two compared gap series (s_1, s_2). For this, both gap series must belong to the same opposition chain and have no intersection of occupied places.

$$ratFunc(s_1, s_2) := \frac{|number\ of\ gaps\ in\ s_1| + |number\ of\ gaps\ in\ s_2|}{|number\ of\ sounds\ of\ the\ opposition\ chain|}$$

In the last step, the sound changes identified in this way are to be typologically checked. The aim of this check is to remove false sound changes caused by gaps that are not based on sound changes. The data of Brown et al. (2013), already presented in Sect. 10.2.4.4, will serve as a basis for this. The frequency of a sound correspondence given there is divided by the frequency of both sounds in the sound-correspondence list. All sound changes with a value below 0.1 are discarded. The list of Brown et al. groups similar sounds together (e.g., *i*, *ɪ*, *y*, and *ʏ* to *i*), which means that no data are available for sound correspondences within a sound group. In this case, a value of 0.25 was applied.

12.2 Evaluation

The result of the described algorithm *before the typological check* is given for German in Tab. 12.C. At a certain point the iteration procedure, no longer leads to relevant results because the algorithm has to remove more and more sounds to get to the "ideal" system. In the test, this point is reached after 15 iterations. After that, the procedure removes entire rows (e.g., nasals in iteration 17–19 or affricates in 24 and 25, see App. E.4), since these consist of too many gaps to remain in the ideal system. Therefore, a distinction is made in the evaluation between with and without threshold (set to the 15th iteration).

Table 12.C. Result of the gap approach with German data after 10 iterations (for more iterations, see App. E.4). The decisive feature of the gap series is marked in italics. In the "Typo" column, all typologically probable sound changes with a value of more than 0.01 are listed. Correct sound changes are marked with *TP*, sound change-related gaps with wrongly identified pre-sound with *dir.*, conditional sound change (with wrong pre-sound) with *cond.*, sounds with single-valued features with *SVF*, and borrowed sounds with *borr*.

	Pre-sound	Post-sound	SC	Typo
1	[affricate, consonant, voiced]*:[alveolar]*	[consonant, voiced, trill]*:[alveolar]*	*d͡z > r (SVF)	
2	[affricate, consonant, voiced]*:[alveolar]*	[consonant, lateral, voiced]*:[alveolar]*	*d͡z > l (SVF)	
3	[affricate, consonant, voiced]*:[palatal]*	[approximant, consonant, voiced]*:[palatal]*	*ɟ͡j > j (SVF)	
4	[consonant, plosive voiced]*:[postalveolar]*	[affricate, consonant, voiced]*:[postalveolar]*	*ɖ > d͡ʒ (borr.)	
5	[rounded, back, long, vowel]*:[open-mid]*	[rounded, back, short, vowel]*:[open-mid]*	*ɔː > ɔ (dir.)	0.25000
6	[rounded, back, long, vowel]*:[open-mid]*	[rounded, short, vowel, front]*:[open-mid]*	*œː > œ (dir.)	0.04334
7	[bilabial, consonant, voiced]*:[fricative]*	[consonant, postalveolar, voiced]*:[fricative]*	*β > ʒ (borr.)	
8	[bilabial, consonant, voiceless]*:[fricative]*	[consonant, labiodental, voiceless]*:[fricative]*	*ɸ > f (TP)	0.09220
9	[bilabial, consonant, voiceless]*:[fricative]*	[consonant, palatal, voiceless]*:[fricative]*	*ɸ > ç (cond.)	
10	[bilabial, consonant, voiceless]*:[fricative]*	[glottal, consonant, voiceless]*:[fricative]*	*ɸ > h (cond.)	0.03571

The presented iteration procedure yields 25 sound changes before reaching the ideal phoneme system; of these, two mergers corresponds with historical sound changes (*ɸ > f and *β > v). Both sound changes are of Germanic origin (cf. Paul et al. 1969:108 fn. 2) and caused that kind of gaps within the phoneme system as it is expected to the theory in the literature. In five cases, a correct gap was detected in the phoneme system, but the correct pre-sound was not determined. These have in common that they include vowels and the original sound could not be determined internally. The gaps are caused by the difference in vowel quality of the opposition *short vowel : long vowel*, which arose in New High German by the synchronic rule that short vowels are open and long vowels are closed (cf. Paul et al. 1969:55, §23.3). While the vowels [a] and [ɛ] deviated from this rule, there is a total of six vowel pairs, four of which were identified. The determination of the pre-sound proved difficult. For the sound [ɔː], both the missing [ɔː] and [o] are possible pre-sounds. Due to the iteration procedure, in which one sound is removed from the phoneme system after each iteration, the “last” opposition pair is no longer identified as such (in this case [ɪ] and [iː], cf. Tab. 12.D).

Table 12.D. The alternants of German long and short vowels (left): Short vowels are open, long vowels are closed. The short vowels [a] and [ɛ] are exceptions. On the right column, the interpretation of the systemic gaps through the algorithm.

Short vowel : long vowel	Identified sound changes
ʊ - uː	*ɪː > uː
ɔ - oː	*ɔː > ɔ and *e > oː
~~a - aː~~	
ɛ - eː and ɛː	
œ - øː	*œː > œ
ʏ - yː	*ɪː > yː
ɪ - iː	

In one case (*ɟ͡j > *j*), the gap arose from the elimination of the oppositional sound: the velar approximant *w* evolved to bilabial [β] around the year 1100 and to labiodental [v] in the 13th century (Paul et al. 1969:100, §76). In this case, the basic hypothesis is wrong. It assumes the reverse case, that a sound (e.g., *f*) without other pairs of the same series (e.g., *labiodental*) arise from a split-off from another series (*bilabial*). If a series consists of only two sounds and one of which disappear (as in the case of the approximants *j* and *w*), it is synchronically indistinguishable from the first case. The sound change *d͡z > *r* and *d͡z > *l* can also be evaluated as real errors. Laterals and trills are represented in the phoneme system with only one sound each, without this being attributable to a sound change. This finding holds for most languages of the world and can be explained by the fact that languages with non-alveolar liquids tend to eliminate them. These sounds, therefore, behave similarly to *j*. In German, *j*, *r,* and *l* are so-called *single-valued features* (Ewen and Hulst 2001:79–

83, cf. *single-relevant features* according to Akamatsu 1988:109). These features only occur once in the phoneme system so that for the sound [l] only the feature *lateral* is “relevant.” Since it is mostly a matter of the manner of articulation, single-valued features have many “gaps” in their series and are therefore highly ranked in the procedure.

Two identified sound changes belong to gaps created by conditional sound change: **Φ* > *ç* (correct: **x* > *ç* before front vowels) and **Φ* > *h* (correct: **x* > *h* before vowels). These are not unconditional sound changes and thus not sound changes “to be achieved.” The determined pre-sound, in this case, can only be false since the underlying assumption of a system-conditioned sound change is false. These cases are identified when a new sound or feature is added to the phoneme system by the conditioned sound change. A determination of these sound changes could be made by the distinctive or phonotactic approach. The situation is similar when the phoneme system borrows sounds. In this test, three identified sound changes included borrowed sounds: **œː* > *ãː*, **β* > *ʒ*, and **ɖ* > *d͡ʒ*.

Since the procedure produces typologically improbable sound changes for the irrelevant gaps in the system, the typological check reliably eliminated the majority of incorrect sound changes (cf. “Typo” in Tab. 12.C). Of the total 25 sound changes, only nine sound changes remained with a value greater than 0.1. These included only two false positives (**Φ* > *h* and **œː* > *ãː*) in addition to the seven correct sound changes. The sound change **œː* > *ãː* hardly occurs anywhere in the world but is nevertheless found in the list because Brown et al. (2013) groups similar sounds together. In this way, the nasal vowel is treated as a simple *a*, which does correspond to *œ*.

12.3 Discussion

The gap approach shows a good result with a precision of 0.7778, but only in two cases, the pre-sound could be determined correctly. The typological validation of the resulting sound changes is an important tool for an appropriate result. Without taking the typological data into account, the approach only achieves a precision of 0.4667 (with threshold) or 0.2222 (without threshold) for German. Of 167 sound changes of the gold standard, only seven could be assigned, resulting in a recall of 0.0324. No phonological rules of the gold standard were identified. The reason for this is that the majority of sound changes and phonological rules are conditioned alternations and therefore not primarily to be achieved by this method. Since the method is the only internal reconstruction method that can capture unconditional sound change, it is of particular interest to IR. Nevertheless, two problematic issues lead to false positives. These problems can only be tackled with typological data.

Table 12.E. Result of the gap approach for German after 15 iterations. Only the categories "sound changes" and "sound changes with wrong pre-sound" are considered true positives.

	Number	**Precision (with threshold)**	**Precision (without threshold)**	**Precision (with typological data)**
Sound changes	2	0.13333	0.08	0.22222
Sound changes with wrong pre-sound	5	0.46667	0.28	0.77778
Sound-change caused false positive	1	-	-	-
Conditioned sound changes	2	-	-	-
Borrowed sounds	3	-	-	-
Single-valued features	2	-	-	-

Firstly, the BASIC HYPOTHESIS IS NOT VALID for some gaps. The procedure assumes that a gap-rich feature has arisen from a series by unconditional sound change. This assumption does not hold for *single-valued features* and *borrowed sounds* that occupy a previously unoccupied position. Similarly, "gap-rich features" can arise from *conditional sound change*. One way to increase precision is to weed out all single-valued features. In the German test data, there are nine of these features: approximant ([j]), near open ([ɐ]), glottal ([h]), lateral ([l]), middle ([ə]), moving backwards ([aʊ̯]), nasalized ([ã]), non-peripheral ([ɔʏ]), and trill ([r]); five of which were identified in the sound changes.

Secondly, the DETERMINATION OF THE PRE-SOUNDS fails in most cases. This proves to be much more difficult since only two correct pre-sounds could be determined. In these cases, an optimal condition was present: the gaps of one series (bilabial) and the gaps of the second one (labiodental) complemented each other; since the labiodental sounds each only occupied one position in the series of the articulation manner, this could be determined as the younger sound according to the basic hypothesis. This situation, however, is likely to account for only some unconditional sound changes. In the case of vowel quality, it was not possible to determine the pre-sound in this way, as several options were available. Some of the sounds identified, such as [t̠], [n̠], and [ɟ͡j], are universally very rare sounds and are absent in most languages. This shows that the use of typological data is inevitable for the gap approach. Phoneme systems tend to have a lot of gaps. Determining the appropriate gap with internal criteria is only possible under optimal conditions.

13. Conclusion

13.1 Summary

13.1.1 Summary of the Theoretical Part

The first part of thesis addressed the issue of Internal Reconstruction from a theoretical point of view and discussed in detail its problems and basic premises as put forward in the literature. The SECOND CHAPTER approached the topic of linguistic reconstruction *per se*. Different views on the question of what linguistic reconstructions should be considered lead to different views on what is methodologically permissible. The linguists which regarded a reconstruction as a historical reality will inevitably reject methods that come to a different conclusion from their preferred method because only one wordform can be historical. Those linguists which see reconstruction as a purely theoretical and non-historical formula have to cope with the question about the purpose of reconstructions. If one assumes that a reconstruction did not exist historically, one can neither assume cognates in sister languages, nor formulate any etymological considerations: the assumption of cognates implicitly presupposes the existence of a historical base form. This thesis followed the middle way; a reconstruction is regarded as an approximation to a historically real wordform and different reconstruction methods might approach this form differently. For the further procedure, this meant that a method does not necessarily have to reconstruct the historically correct wordform in order to find acceptance as a method. Rather, the precision of the results gives an indication of the reliability of the method.

The thesis assumed four different reconstruction methods: Comparative Method, Internal Reconstruction, External Method, and Philological Method. As outlined, these four methods followed the same reconstruction scheme containing a presuppositional phase, an inductive procedure, and inferring methods. The major difference between these methods is their different basic hypotheses. From these, the results and the reliability of the method, are ultimately derived. The basic hypothesis of the Comparative Method is not general valid, but it proved to be more reliable compared to the basic hypothesis of IR. The basic hypothesis of IR assumes that two synchronic allomorphs are traced back to the same historical pre-form. For suppletion and analogy, this basic assumption does not hold and, thus, reduce the reliability of the method. For the Comparative Method, there are similar confounding factors, such as loanwords or neologisms, but these are easier to identify as such. Since the relative unreliability of IR is inherent to its basic hypothesis, there is no possibility to increase the reliability of IR, not even with computational approaches with larger data sets.

In the THIRD CHAPTER, it was shown that this directs the scope of IR to the peripheral area of historical linguistics where the Comparative Method is not fruitful. This includes isolated languages, pre-proto-languages, the reconstruction of more recent language stages, and the establishment of relative chronologies of sound changes and wordforms. Since other reconstruction methods are not or only with difficulty applicable for these fields, the results of IR are hardly to be evaluated. Besides a definition of IR via the basic hypothesis (i.e., the *structuralist definition*), a more general definition is also in use and works with the singularity principle: all means are allowed for IR as long as they can be derived synchronically and without dialectal or phylogenetic information. This thesis followed the latter definition and allows the application of linguistic knowledge. This definition is important for automated IR because the structuralist definition may be associated with scientific stagnation for CHL. Using the structuralist definition, it would be neither expected to produce a more accurate result due to the larger amount of data, nor to be a support for the study of poorly researched languages because determining the older of two alternating wordforms is usually not an effort for linguists. The change to other internal reconstruction methods is also accompanied by a change of the basic hypothesis.

This leads to the definition of different sub-methods of IR, which are presented in CHAPTER 4. The classical IR is the *morphophonemic method*. Depending on the source data, three types can be distinguished. The paradigmatic approach works with alternations within a paradigm, whereas the derivational approach with means of productive word formation. For most part, allomorphs in this domain form internal cognates. Their alternations are exclusively *conditioned*. The semantic method is the third type of morphophonemic method and assumes cognacy of semantically related words that do not reveal productive word-formation affixes. Here, the hypothesis of etymological relatedness is far more often incorrect. Another method of IR that is only peripherally picked out as a central theme in the literature is the *distributional method*. It starts from the assumption that significant deviations in the distribution of phonotagms in a language are the result of sound change. Namely, an absence or marginality of a phoneme cluster or sound are assumed to be of historical origin, while a disproportionate accumulation of a phoneme cluster indicates an emergence or increase due to sound change. This method is often supplemented by the additional assumption that, formerly, there was a symmetrical phoneme system which had been broken up by sound change. The general validity of the distributional basic hypotheses is theoretically difficult to maintain, but they can be used to reconstruct different types of sound change, such as unconditioned merger. For CHL, this method is of particular interest since it can be implemented with low resource, and detailed mathematical analysis may provide insights that are not apparent in manual reconstruction. A method similar to the morphophonemic method is the so-called *pattern reconstruction*. Instead of assuming a relationship between two morphs, it assumes a relationship between two structural patterns. For example, two patterns with the same grammatical function, such as *CV̄* and *CVC* can thus be traced back to a pre-pattern **CVC*. The main difference with the morphophonemic method is that the abstraction of the individual sound correspondences occurs last in the morphophonemic method (e.g., the alternations *ā:ah* and *ī:ih* to **h* > 0 / V>V̄_), while pattern reconstruction takes this step first (*ā, ī* to *V̄* and *a, i* to *V*, then *V̄:Vh* and **h* > 0 / V>V̄_). The *onomatopoetic method* of IR assumes that a wordform was originally formed as onomatopoeia and that phonetic deviations between the wordform and the "original form" arose through sound change. One problem for the computational implementation of this method is that the original phonetic form is not attested and must be constructed.

The four mentioned methods of IR serve to reconstruct *pre*-lingual wordforms and *pre*-sounds. Depending on the goal, one speaks of wordform or sound-change reconstruction. Furthermore, there is the possibility of internal reconstructions of semantics, syntactic structures, or grammatical categories, which can be summarized under the term "grammatical IR." One of its most popular method is the *archaistic method*. Out of two forms (e.g., two semantics of a word), it considers the form as the older one that is synchronically considered to be more archaic or unproductive. Its computational implementation is difficult. Although some linguistic resources provide the information on obsolete usage, this usage reference is mostly the result of diachronic research. Other grammatical methods work with homonymy, suppletive forms or special principles and assumptions. These have in common that they are often to specialized to make a computational approach worthwhile, or the linguistic resources needed are not available. Missing resources are also the problem for *typological IR*. This method attempts to derive pan-linguistic rules from empirical data, which, in turn, can be integrated into the reconstruction process. However, the thesis viewed cross-linguistic rules less as reconstruction methods and preferred to regard them as inferring methods.

The *decompositional method*, as it is found in the pre-linguistic and occasionally in modern literature, is not to be regarded as a legitimate reconstruction method. It assumes that a word is etymologically a compound that lost its original form through sound change. An etymologist may interpret any compositional elements into the word in question. The more sound or semantic change and word-formation affixes he hypothesizes, the more arbitrary is the result. This danger is equally present for internal reconstructions of agglutinative languages because, in these languages, a single sound functionally may also represent an affix. For these methods, it is therefore necessary to assume as little linguistic change as possible.

In the scientific history, IR has repeatedly been redefined and methodologically adapted (as shown in CHAPTER 5). Its definition and its goals were always a reflection of the theories that were dominant at that time. Formative work was done in particular by the structuralist Hoenigswald, with whom a concentration on the morphophonemic method took place. However, this method was increasingly criticized in later times for differing little from synchronic phoneme analysis. Therefore, different new approaches were developed — often with little acceptance. In fact, most of the criticisms of IR, as listed in CHAPTER 6, relate to the morphophonemic method. These include the unusual regularity of the reconstructed languages and the temporal indeterminacy of the results. It should be noted, however, that the term "pre-lingual" does not mean a uniform language stage, and therefore a regular language stage cannot be assumed. Other criticisms, such as the invalidity of the basic hypothesis in the case of suppletion, the incompleteness of the reconstructed language and the speculative nature of the reconstructions, are, to some extent, common to all reconstruction methods. The non-morphophonemic methods are frequently considered too subjective (e.g., the decision which grammatical form is considered older may differ from linguist to linguist). One advantage of a computational IR is that it may objectify subjectivity. I see the future focus of computational IR primarily in the optimization and expansion of these non-morphophonemic methods. These methods are of particular interest for proto-languages since alternations in proto-languages are reconstructed by the linguists. A computational morphophonemic reconstruction on proto-languages would only reflect the reconstructing linguist's idea of the pre-proto-language.

13.1.2 Summary of the Computational Part

The second part of the thesis described the machine implementation of internal-reconstruction methods, choosing the morphophonemic and distributional methods. The morphophonemic method was implemented in its three variants: the paradigmatic, derivational, and semantic approach, the distributional method in a phonotactic, distinctive, and gap-systematic approach. Unlike most automatic models of the Comparative Method, which mostly reconstruct word-forms (*Ancestral State Reconstruction*, ASR), this work aimed at the reconstruction of sound changes. Word-form reconstructions with IR are usually associated with difficulties, such as the already mentioned anachronism of individual sounds of a reconstructed word or the problem of determining the original sound (especially with distributional methods). The FIRST GOAL of the work was to automate the reconstruction methods themselves. For the semantic and distributional methods, this also meant a formal description of the algorithm, which is missing in the literature so far. In this respect, the thesis also saw itself as a pioneering work aimed at developing and optimizing innovative methods. The SECOND GOAL of the thesis was to evaluate IR as a reconstruction method. This means, how much of the historically-known sound changes could be reconstructed using IR, what the error rate was, and whether IR tended to reconstruct phonological rules rather than sound changes. It should be noted, however, that errors in the source data or the implementation are expected to lead to a different result from a manually performed IR. In this respect, the work was intended as a study to determine and answer questions about the performance of IR. The test languages used in the thesis were German as a historically well-attested language and Proto-Indo-European as a reconstructed proto-language. German is attested since the early Middle Ages and allows an evaluation with three language stages. The gold standard is mainly the result of the Comparative and Philological Method. The application to a proto-language was mainly to compare the results of automated IR with those of manual IR since IR is the only reconstruction method here. In the absence of a gold standard for Pre-Proto-Indo-European, the results of the automation were compared to the results that had already been proposed in the literature.

For the implementation of the *paradigmatic* and *derivational approach* of the MORPHOPHONEMIC METHOD, data from the online dictionary Wiktionary served as source data. This dictionary provided IPA transcriptions as well as paradigms and derivations. In the first step, the transcribed wordforms were reduced to the root and all allomorphs of the same word (e.g., *Hilfe* 'help, aid' → *helf, hilf, half, holf*) were aligned. A Levenshtein distance matrix was used for this purpose, which determined the sound sequence alignment with the lowest number of substitutions (i.e., *Levenshtein distance*) as the basis for sound correspondences (e.g., *h:h, e:a, l:l, f:f*). The morphophonemic method was implemented as a threshold-based approach: in order to separate correct from incorrect sound correspondences, these were evaluated using the Jaccard index as relevance measure. Incorrect alignments due to suppletive forms become statistically insignificant and did not interfere with further processing. Conversely, very rare sound alternations that may reflect ancient sound change were repressed. To determine the conditions of sound change, a relevance measure (*conDiff*) for each context was introduced. This compared the relative frequency of a context in all wordforms containing the respective alternation. Possible contexts were the environments *_X, X_, X_Y,* and the occurrence of a sound *X* in any position of the wordform. In order to optimize the result of the relevance measure, a significance test (*Chi-squared four-fold test*) was multiplied by the relevance measure as an indicator function. In the last step, the sound correspondences were abstracted to sound laws. For this purpose, all sounds were decomposed into their distinctive features and the sound pairs were converted into *feature pairs*. A sound correspondence with a-many distinctive features of the first sound and b-many features of the second sound became $a \cdot b$ feature pairs. Each feature pair inherited the alternating wordforms. In this way, the procedure for determining the relevance measure for sound correspondences could be applied to feature pairs. Similar to the sound pairs, the conditions (i.e., *X_, _X, X_Y*, and *X*) were decomposed into all possible distinctive features and evaluated in the same way. To reduce a large number of duplications, contexts that were merely a subset of another context have been eliminated. Since many sound laws consist of exclusively concrete sounds, dominant sound pairs were also allowed as sound laws in the final evaluation if they were dominant in their feature pair. The result of the evaluation can be found in the following section.

Ontologies were used to automate the SEMANTIC APPROACH. Ontologies are semantic networks in which concepts are connected by their semantic relation to each other. For each concept in question, a bag of words was created containing all hyponyms, co-hyponyms, and hypernyms of the concept. To identify the internal cognates within this bag of words, the phonetic similarity of each word to the others was determined using a weighted sequence alignment. The probability of a sound change between two sounds was defined using a PMI score and each word pair in the bag of words was aligned by an iterative Needleman-Wunsch algorithm. The procedure yielded a similarity score for each word pair, with a positive value indicating cognacy. Although this method was able to distinguish true internal cognates from non-cognate words in individual concepts, the high error rate did not allow for the extraction of sound changes from these lists. Overall, only 14.71% of the word pairs were cognates (in comparison: for the other morphophonemic methods, this value was over 99.9%). Rather, the semantic approach seems suitable for computer-aided detection of potential internal cognates. Another possible application is the detection of word formatives that have become unproductive. Word formation patterns were searched in the list of potential cognates and the function of the identified word formative were semantically interpreted with the help of ontological relations. This procedure was tested for verbs of Proto-Indo-European, which contain so-called root extensions as an unproductive word formative affix. The test was based on German sound-change probabilities and an ontology designed for German, which visibly affected the result negatively. Three “morphemes” *k* (for hyponyms), d^h, and *i̯* (each for hypernymy) were identified, which were already known as root extensions and infixes, respectively, at least for individual verb pairs.

The distributional method was implemented in this thesis in three possible ways. The so-called PHONOTACTIC METHOD tries to infer historical sound changes via the distribution of phonotagms in a language. The aim of this work was to formalize the method and to test its reconstructive capabilities. Both an IPA-transcribed corpus and word lists were used as source data. Corpora have the advantage that the proportion of loanwords is lower, while the advantage of word lists is the large number of different words. For each sound, a feature vector was created which contained the feature pattern *_X* and *X_* as vector scalars. A trigram pattern *X_Y* proved inappropriate due to sparse-data problems. The occurrences of the features reflected, in normalized and weighted form, how often the sound in that position occurred with another sound. The conditions of a sound change resulted from the difference for a feature *_X* in the two sound vectors, $\vec{a}$ and $\vec{b}$. A high positive value indicated a condition for the sound *a*, a high negative value indicated a condition for *b*. The difficulty of the method lied in the determination of the sound pairs. To determine a relevant sound, it was looked for features in the vectors that were outliers compared to values in similar vectors. If a context occurred far more frequently in a sound than in similar sounds, this indicated a sound change in that context. To identify the complementary sound, a pure-phonotactic, four phonetic and two cross-linguistic approaches were tested in this work. A pure-phonotactic approach tries to find a similar sound that hardly occurs in relevant contexts because it has disappeared due to sound change. Finding the complementary sound seems to be barely possible with purely phonotactic approaches; the historically correct complementary sound is phonotactically too similar in most cases. Four phonetic approaches and two further cross-linguistic approaches have been tried to determine the complementary sound. Some of them have been proposed in other contexts of IR in the literature: The sound with complementary distribution may be determined as a complementary sound of sound *a*, which:

- has at least one distinctive feature in common with sound *a*,
- is articulatory most similar to sound *a*,
- is most similar to sound *a* without the dominant feature of the conditions,
- belongs to the same Dolgopolsky class,
- strongly differs from its phonotactics in typologically similar languages, or
- empirically frequently appears with the sound *a* in sound changes.

The third method, which took conditions into account, and the empirical variant were able to determine the pre-sound in about half of the sound changes tested and performed best. However, the former is only suitable when an assimilation process is present, and the latter is unsuitable for atypical sound changes. The fifth approach compared the sound vectors to "ideal vectors," which reflected the average occurrence of a sound at the contexts in question in phonotactically similar languages (*sub-typological approach*). If its occurrence in a context deviated significantly from this, this indicated a sound change. For the test, 60,351 IPA-transcribed words from about 300 languages were used and sounds missing in German were assigned to a similar German sound. However, a determination of the complementary sound could not be achieved in this way, since, in most cases, the resulting "ideal corpus of phonotactically similar languages" was phonotactically far more dissimilar than would be suitable for comparison. Replacing a lemma list with a morpheme list had a negative effect on the result. For the application case with Proto-Indo-European, a word list of verbal roots from LIV was used. The evaluation of the identified sound changes after ten iterations showed that only two phonological rules were mapped. The other cases were salient phonotagms which are universal (e.g., phonotactical contrast between consonants and vowels) or became "relevant" due to reconstruction.

The phonotactic method is generally suitable for reconstructing conditional sound change, but two problems speak against a practical application. On the one hand, there is the problem of dealing with irrelevant phonotagms, which made up a large part of the result, and, on the other hand, there is the problem of correctly determining the correct sound pair. The pre-sound of a sound changes and the conditions were frequently interchanged, making the originating sound hardly determinable.

The DISTINCTIVE APPROACH tries to implement distributional approaches that involved the relative phonetic frequency and the symmetry of the phoneme system. It gets its information from minimal pairs whose number may change due to sound change (especially due to merging). For this purpose, a transcribed morpheme list of German was used as source data. For each sound, a morpheme set with this sound was created and all occurrences were replaced by a placeholder. The intersection of two morpheme sets resulted in the number of minimal pairs. From the size of the intersection, statements about their sound classes and allophony could be derived, which were illustrated by clustering the sounds. The distribution of the sounds within the clusters indicates historical sound changes, but conclusions about the concrete changes were difficult. In order to determine sound correspondences, the sound pairs were evaluated by means of two relevance measures, which considered the phonetic frequency and the phonotactic similarity or the empirical probability of the sound correspondence. Again, to reduce the number of duplications, an iteration procedure was used in which any sound correspondence containing a sound of the best sound pair was removed. The possible conditions of a sound change were #_, #C_, _#, and _C#. The result of the procedure is a list of 13 sound changes, seven of which could be assigned to a historical sound change. Similar to the phonotactic method, the correct determination of the complementary sound and the conditions is still an open issue. The application of the method to Proto-Indo-European could identify Siebs's law and the so-called *b*-gap.

The GAP APPROACH tries to draw a conclusion about historical merging from "gaps" within the phoneme system. For this purpose, all distinctive features of the test language were determined and an "ideal phoneme system" was created, in which all possible combinations of these features were considered. This was compared with the actual phoneme system and an iterative process was implemented to close these gaps by assuming as few sound changes as possible. After about 15 iterations, a point was reached where the algorithm no longer produced any reasonable sound changes, as it had to remove more and more sounds to reach an "ideal system." For German, the procedure only identified two correct sound changes. In five other cases, the correct pre-sound was not determined because the original sound could not be reconstructed internally in this way. An obstacle of this method was borrowed sounds and so-called "single-valued features," since these contradict the basic assumption of the method. The number of false positives could significantly be reduced using empirical probabilities.

Table 13.A. Summary of all implemented approaches. PR/SC indicates whether the method is more suitable for phonological rules (PR) or sound changes (SC). The number indicates the number of single sound changes detected with the method for German data.

	Method	**Basic hypothesis**	**Reconstructable SC**	**PR/ SC**	**Detected SC**
morphological	paradigmatic	Cognacy of paradigmatic allomorphs	Conditioned sound change	PR	17
	derivational	Cognacy of derivational allomorphs	Conditioned sound change	PR	18
	semantic	Cognacy of semantic similar words	Conditioned sound change	SC	-
distributional	phonotactic	Phonotagms reflect historical sound changes	Conditioned sound change (without loss)	PR	7
	distinctive	Distinctiveness of phonemes reflects sound changes	Conditioned sound change (without loss)	PR	7
	gap	Phonemic gaps result from sound loss	Unconditioned loss and merger	SC	7

13.2 Evaluation

13.2.1 How Many Sound Pairs Can Be Assigned to a Historical Sound Change?

The first question that should be answered by the evaluation is how many of the identified sound pairs of the German test dataset actually could be assigned to a historical sound change. For this, a MATCH RATE was determined which assigns each machine-identified sound pair to a sound change *or* phonological rule from a gold standard, regardless of the correct change direction or conditions. Duplicate sound changes, which may result from free variants or inaccurate determinations of the sound change, were excluded from the evaluation. Among the first 100 best-evaluated single-sound pairs, the derivational approach found 17 of 26 single phonological rules and 18 of 167 sound changes since Early Germanic. As Tab. 13.B shows, the results with the distributional methods are significantly lower (five or three phonological rules and seven sound changes). In the phonotactic method, the best "precision with doublets" (i.e., the same sound changes and phonological rules with a different condition is considered "true positive") is 0.4211 (for phonological rules) and 0.8 (for sound changes), respectively, and is thus comparable to the morphophonemic methods. However, the threshold and number of iterations play an important role here. In the paradigmatic approach, the majority of the identified sound pairs with a Jaccard value of more than 0.64 can be assigned to a historical sound change or a phonological rule. Most sound pairs below this value reflect morphological alternations, especially *ablauting* vowels and alternations from multiple alternation paradigms. Erroneous alternations (e.g., from transcription errors) occur more frequently for pairs with a Jaccard value less than 0.37. This results in different F-scores for the methods depending on the threshold. The phonological rules tend to reach the highest F-score with a higher threshold than sound change, since the sound pairs identified first represent mainly phonological rules. For the paradigmatic, derivational, phonotactic, and distinctive method, the F-score values for phonological rules are significantly higher than for sound change. This fact suggests that IR tends to reconstruct rather than phonological rules — a tendency that is particularly true for the distinctive and morphophonemic methods. The situation differs from the gap approach, which is unsuitable for reconstructing phonological rules because it assumes a change that has caused a gap in the phoneme system. This assumption is hardly true for any phonological rule. However, despite the low F-score of distributional methods, this is precisely where they have potential. Further research and optimization are imperative here to make them suitable for practical applications.

Table 13.B. Result of the evaluation (match rate) of five different IR methods with German test data. The evaluation considered the first 100 top-rated individual sound pairs and all German sound changes starting from Early Germanic. Bracketed numbers indicate the positioning of the sound pair when the highest F-score is reached. The numbers in italic cells indicate the values for abstracted sound changes.

		PR	**SC**	**Highest F-score (PR)**	**Highest F-score (SC)**
Paradigmatic	single	12	17	0.32258 (6)	0.11184 (83)
	abstract	*5*	*9*	*0.5 (3)*	*0.13636 (33)*
Derivational	single	17	18	0.36066 (39)	0.12167 (39)
	abstract	*6*	*9*	*0.46154 (6)*	*0.29032 (23)*
Phonotactic		5	7	0.17021 (35)	0.06178 (55)
Distinctive		3	7	0.15385	0.06114
Gap		0	7	0.0	0.05809

Tab. 13.B also gives the results for the transformed sound changes and phonological rules, which are similar to those of the single-sound evaluation. Because the transformed sound pairs consist entirely of distinctive features, the result depends on the individual sound changes. While some pairs matched the historical sound changes well (e.g., the feature pair *back_vowel-front_vowel* and the German *umlaut*), distinctive features proved rather inappropriate for other sound changes (e.g., *vibrant-central_vowel* for *ʁ* > *ɐ*) or not enough (e.g., *long_vowel-short_vowel* for V[+stressed,-long] > V[+stressed,+long]). Only 17.3914% of the individual sound pairs covered by the identified feature pair belonged to a historical sound change using the paradigmatic method

(20.6335% for phonological rules). With the derivational method, the precision value was significantly lower at 5.1515% (5.217% for phonological rules).

13.2.2 How Often Could the Direction of Sound Change Be Correctly Determined?

The direction of a sound change or phonological rule was evaluated by calculating the rate of sound pairs with the correct direction of all matched sound pairs. Three approaches were used to determine the direction for the morphophonemic methods: the rule of the phonetically most plausible direction, the rule of articulatory proximity to the condition, and the rule of the restricted sound. As can be seen from Tab. 13.C, the rule of reconstructing the sound-change directions using the condition was able to provide the best results, followed by the rule of the restricted sound. The plausibility rule mostly fails because both sound-change direction are plausible and therefore no direction could be determined. However, if a statement is possible, it will make a valid direction reconstruction.

Table 13.C. Rate of the correct direction of sound changes for three different approaches. The reconstructions were done with the paradigmatic and derivational approaches and the results were compared with the gold standard of both phonological-rules (PR) and sound-changes (SC). The values in parentheses indicate the results of the method with abstract sound changes.

	Paradig. Approach		**Derivational Approach**	
	PR	*SC*	*PR*	*SC*
plausibility rule	0.16667 (0.2)	0.05882 (0.22222)	0.08333 (0.16667)	0.08333 (0.11111)
using the conditions	*0.41667* (0.4)	*0.47059* (0.33333)	*0.41667* (0.5)	0.5 (0.55556)
rule of restricted sounds	0.33333 (0.4)	*0.47059* (0.44444)	0.16667 (0.66667)	*0.75* (0.7778)

With the phonotactic approach, the determination of the original sound was not readily possible. Instead, the problem arose here that the “resulting sound” and “condition” had to be determined. The method always identified the rarer sound of a bigram phonotagm as the condition of a sound change or phonological rule and is therefore hardly suitable for determining the direction. Due to the small number of identified sound changes and phonological rules, no significance can be inferred from the rates in Tab. 13.D. Nevertheless, the values indicate that the hit rate must be close to the random probability of 0.5. Therefore, the determination of the changed direction does not seem to be possible internally.

Table 13.D. The correct change direction among the correct sound pairs after 100 iterations. The approach of using the conditions for determining the complementary sound was applied for the paradigmatic and derivational method. In the "phonotactic" row, the correct determination of "resulting sound" and "condition" was measured.

		Phonological Rules		**Sound Changes**	
		absolute	*precision*	*absolute*	*precision*
paradigmatic	single	5 of 12	0.41667	8 of 17	0.47059
	abstract	2 of 5	0.4	3 of 9	0.33333
derivational	single	5 of 17	0.41667	6 of 19	0.5
	abstract	3 of 6	0.5	5 of 9	0.55556
phonotactic		0 of 5	0.0	0 of 7	0.0
distinctive		2 of 3	0.66667	4 of 7	0.57143
gap approach		0 of 0	0.0	2 of 7	0.28571

13.2.3 How Often Could the Conditions of Sound Change Be Correctly Determined?

For the evaluation of the conditions, a threshold value of 0.9 was introduced, which an identified condition must reach to be considered "positive." True positives were those that represented an actual triggering condition of a sound change or phonological rule. Thus, a correct condition _V was satisfied by the conditions r_a, _e, or _i. Phonological rules achieved precision up to 85.648% (paradigmatic, single-sounds); for sound changes, this was lower. If exclusively unproductive sound changes were taken into account, the precision for sound-change conditions would be only 31.25% (paradigmatic, single-sounds) and 0.0% (derivational, single-sounds). A finding that suggests that this method is more likely to map phonological rules. The conditions of unproductive sound changes are lost by secondary developments or disappear below the threshold rate of 0.9, where they can hardly be distinguished from the “false” conditions. Most of the true positives were duplications of the same historical conditions. The doubling rate in Tab. 13.E reflects the percentage of duplications among the true positives. It was lower in the derivational approach because there are more analogical levelling and therefore many conditions do not reach the threshold. For the distinctive approach, only a few conditions were possible to reconstruct at all and therefore cannot be directly compared. With this approach, the precision was 0.7143. The gap approach was designed to reconstruct unconditioned merging and did not reconstruct any conditions.

Table 13.E. Precision values and doubling rate of the morphophonemic methods. The threshold was set to 0.9. In parentheses, the absolute numbers of true positives are given. For the phonotactic approach, the evaluation of the conditions was combined with the evaluation of the complementary sound (see Tab. 13.D).

		Phonological Rules		**Sound Changes**	
		precision	*doubling rate*	*precision*	*doubling rate*
paradigmatic	single	0.85648 (185)	0.83520	0.78919 (146)	0.72777
	abstract	0.54008 (128)	0.90871	0.44948 (129)	0.72697
derivational	single	0.48718 (38)	0.45152	0.57143 (36)	0.55185
	abstract	0.22951 (84)	0.68031	0.19355 (84)	0.68031

13.2.4 How Old Are the Internally Reconstructed Sound Changes?

Among the first 100 sound pairs, the paradigmatic approach found 17 sound changes, two of which could be attributed to the Pre-Old High German, nine to the Pre-Middle High German, and six to the Pre-New High German period. Of the 18 sound pairs of the derivational approach, one was Pre-Old High German, eleven were Pre-Middle High German, and six were Pre-New High German sound changes. Due to the *umlaut*, which took place before Middle High German and is still recognizable in the synchronic stage, this phase reached the peak of these three periods. With abstract sound pairs summarizing all *umlauts*, the peak is in the Pre-New High German period. The lower precision values in Tab. 13.E for the derivational approach has mainly to do with the changing stress positions in compounds, which were found to be more relevant and thus postponed real sound changes. Recall, which measures how many historical sound changes it might has missed in a period, falls significantly with each older period. The assumption that the older a sound change is, the less likely it is to be reconstructed internally seems to be confirmed. The assumption that the reconstructed sound changes must rather be assigned to the youngest period, on the other hand, cannot be confirmed. On the contrary, the Proto-Indo-European *ablaut* appeared very frequently, so that no statement can be made about the time of the reconstructed sound changes. However, it is noticeable that sound changes that have no or only a few conditions above a threshold of 0.9 are old sound changes. A temporal estimation can therefore be made in this way.

Table 13.F. Temporal evaluation of the identified sound changes. The sound changes are grouped into three periods: since Early Germanic (Pre-Old High German), after Old High German (Pre-Middle High German), and after Middle High German (Pre-New High German). In each cell, the value with the best F-score is given (see App. B.3–B.5 and C.3–C.5). The older periods include the younger periods. "s" stands for the single-sound evaluation; "a" for the sound pairs transformed to abstract sound laws.

		Pre-Old High German		**Pre-Middle High German**		**Pre-New High German**	
		precision	*recall*	*precision*	*recall*	*precision*	*recall*
Paradig.	s	0.21918	0.09581	0.19178	0.125	0.06818	0.07870
	a	0.31035	0.08738	0.5	0.12245	0.33333	0.10526
Derivat.	s	0.34043	0.07407	0.45455	0.13393	0.15152	0.06758
	a	0.45	0.21429	0.35	0.15909	0.25	0.10870

The phonotactic and distinctive approaches show similar results like the morphophonemic methods. Most sound changes identified with those methods originate from the younger periods, as well, and are often also productive phonological rules at the same time. Only the gap approach deviates here. All seven recognized sound changes of the gap approach are unproductive and therefore of special interest, especially since all seven sound changes were not recognized by any of the other methods.

13.2.5 Is It Rather Phonological Rules or Sound Change That Is Reconstructed?

In all methods, the number of identified sound changes was higher than the number of phonological rules (cf. Tab. 13.D). Nevertheless, it could be observed that most of the identified sound changes at the same time represented synchronic phonological rules. This observation and the F-score values speak for *a reconstruction of phonological rules* instead of sound changes. This is also indicated by the reconstruction of conditions, which gave better results for phonological rules (Tab. 13.E). Conditions of sound changes may be covered by secondary developments, which makes reconstruction much more difficult. An exception is the gap approach that indeed could not reconstruct any phonological rule but some sound changes. This can be explained by the fact that gaps in phoneme systems are less often caused by synchronic rules and that a gap can be transmitted for a very long time (e.g., German labiodental consonants).

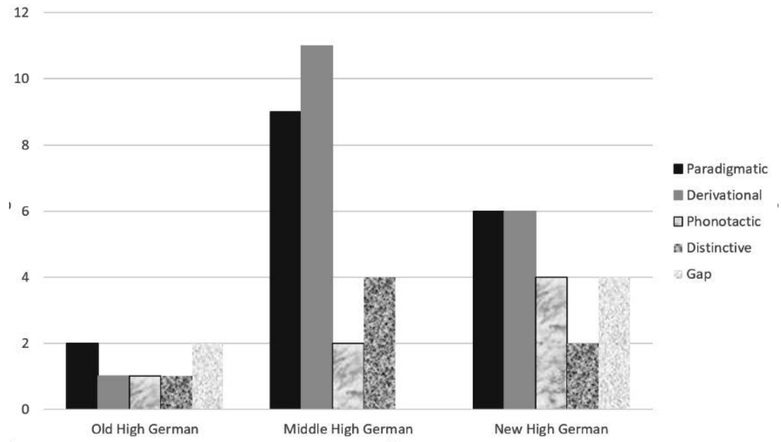

Figure 13.1: Number of identified sound changes and their assignment to a period: Paradigmatic (single sounds, first bar): 2 Pre-Old High German, 9 Pre-Middle High German, and 6 Pre-New High German; paradigmatic (abstract): 3, 2, 4; derivational (single sounds, second bar): 1, 11, 6; derivational (abstract): 2, 2, 5; phonotactic (third bar): 2, 0, 5; distinctive (fourth bar): 1,4,2; gap approach (fifth bar): 3, 0, 4.

Nevertheless, the reconstruction of sound changes that do not have a synchronic equivalent has succeeded several times. Especially old alternations can only be observed with the morphophonemic method in single paradigms or derivations. Quantitative approaches do not offer any advantage here compared to manual IR and do not suggest any optimization possibilities.

13.2.6 Which Resources Are Suitable for Automated Internal Reconstruction?

In this work, different source materials were used for the different reconstruction methods. The paradigmatic and derivational approaches drew their information from an online dictionary. Due to their need for paradigms and word formatives, such dictionaries are a suitable tool for these methods. However, since these were semi-automatically analyzed, errors or inconsistent representations also occurred; but for the most part, they were not significant. For the semantic method, ontologies were used. The semantic relations extracted from there contained many non-cognates or real derivations, while the desired old cognates were hardly found among them. For Proto-Indo-European, a German verb ontology was used, since no resources are available for proto-languages. By necessity, incorrect verb relations occurred due to translation, so only ontologies of the same language may be recommended for this method.

For the phonotactic method, a text corpus, a lemma list, and a morpheme list were used. Using the text corpus, word repetitions and inflectional morphemes distorted the actual frequencies. The lemma list, on the other hand, contained many loanwords, resulting in many borrowed phonotagms, and phonotagms from morpheme boundaries. The use of a morpheme list did not compensate for this disadvantage. For the distinctive method, a minimal-pair list derived from a morpheme list was used. From the viewpoint of resource acquisition, the distributional methods are, therefore, particularly suitable for proto-languages and poorly researched languages.

13.3 Future Works

To conclude this thesis, an overview is given to illustrate where I see the focus of future works on the topic of automated IR. The morphophonemic methods require paradigms or derivations as source data, which are extracted by the tool *eo ipso* from the German Wiktionary. In this respect, the tool may be extended by other sources or self-generated lists of paradigms and derivations. This step is necessary since the focus of the morphophonemic method will rather be on poorly researched languages. However, it remains an open question whether the coarse-grained statistical approach is optimal for this method, or whether it should purposefully look for irregularities in paradigms and derivations instead. In many cases, a relevant alternation is only attested in a single paradigm, and, therefore, is statistically inconspicuous.

The semantic approach is less suitable for determining full cognates because internal cognates are mostly partial cognates. The future focus will be less on phonological reconstruction but rather on the reconstruction of old morphemes (as represented in Sect. 9.5) or semantics (as in List 2019). Through the usage of cognate detection or ontologies that map relations in an appropriate way, the method may be well optimized.

For the distinctive approach, a cross-linguistic study is worthwhile that attempts to derive propositions from the synchronic clusters and their historical sound changes. In this way, an interpretation of the clusters should be facilitated.

Additional studies are needed to automate other internal reconstruction methods. This applies, for instance, to the onomatopoetic method (Sect. 4.2.3) or Hoenigswald's hypothesis of homonymy (Sect. 4.3.6), which entails that, among polysemous lexemes of a language, individual sounds appear more frequent compared to the rest of the vocabulary. This could be interpreted as an indication of a historical merger.

Bibliography

Abels, Sven & Hahn, Axel (2005): Pre-Processing Text for Web Information Retrieval Purposes by Splitting Compounds into Their Morphemes. In: *First Workshop on Open Source Web Information Retrieval,* [Compiègne 2005] (ed. by Michel Beigbeder and Wai Gen Yee). Compiègne/France. pp. 7–10.

Ackermann, Katsiaryna (2018): Root Variation vs. Root Extension – a Putative -*i*-Infix? In: *Historische Sprachforschung* 131(1). pp. 98–144.

Adams, Douglas Quentin (1985): Internal Reconstruction in Mutsun Morphology. In: *International Journal of American Linguistics* 51(4). pp. 329–331.

Adelaar, Karl Alexander (1983): Malay Consonant Harmony: An Internal Reconstruction. In: *Linguistic Studies of the Languages of Indonesia* 16. pp. 57–67.

Adelaar, Willem (2013): Searching for Undetected Genetic Links between the Languages of South America. In: *Historical Linguistics 2011. Selected Papers From the 20th International Conference on Historical Linguistics* [Osaka 2011] (ed. by Ritsuko Kikusawa and Lawrence Andrew Reid). Amsterdam and Philadelphia. pp. 115–128.

Akamatsu, Tsutomu (1988): *The Theory of Neutralization and the Archiphoneme in Functional Phonology* (= Current Issues in Linguistic Theory 43). Amsterdam and Philadelphia.

Andersen, Henning (1973): Abductive and Deductive Change. In: *Language* 49. pp. 765–793.

Andersen, Torben (1989): The Päri Vowel System with an Internal Reconstruction of its Historical Development. In: *Journal of African Languages and Linguistics* 11(1). pp. 1–20.

Anttila, Raimo (1968): The Relation between Internal Reconstruction and the Comparative Method. In: *Ural-altaische Jahrbücher* 40. pp. 159–173.

Anttila, Raimo (1972): *An Introduction to Historical and Comparative Linguistics*. New York City.

Anttila, Raimo (1973): Internal Reconstruction and Finno-Ugric (Finnish). In: *Current Trends in Linguistics* 11*: Diachronic, Areal, and Typological Linguistics* (ed. by Thomas Albert Sebeok). The Hague. pp. 317–353.

Anttila, Raimo (1989): *Historical and Comparative Linguistics* (= Current Issues in Linguistic Theory 6). 2nd edition. Amsterdam and Philadelphia.

Ascoli, Graziadio Isaia (1872): *Vorlesungen über die vergleichende Lautlehre des Sanskrit, des Griechischen und des Lateinischen*. Translation of *Lezioni di fonología comparata del sanscrito, del greco e del latino,* Florence 1870 (translated by Johann Bazzigher and Heinrich Schweizer-Sidler). Halle.

Austerlitz, Robert (1978): On the Use and Abuse of Internal Reconstruction. In: *Ural-Altaische Jahrbücher* 50. pp. 1–6.

Austerlitz, Robert (1982): Gilyak Internal Reconstruction, 1: Seven Etyma. In: *Folia Slavica* 5. pp. 81–88.

Austerlitz, Robert (1983): On the Penumbra of Questions Surrounding the Internal Reconstruction of Gilyak. In: *Research on Language & Social Interaction* 16(3–4). pp. 75–87.

Austerlitz, Robert (1984): Gilyak Internal Reconstruction, 2: Iron and Questions Related to Metallurgy. In: *Folia Slavica* 7(1–2). pp. 38–48.

Austerlitz, Robert (1986): Contrasting Fact with Fiction: The Common Denominator in Internal Reconstruction, with a Bibliography. In: *Linguistics across Historical and Geographical Boundaries. In Honour of Jacek Fisiak on the Occasion of His Fiftieth Birthday*, vol. 1*: Linguistic Theory and Historical Linguistics* (ed. by Dieter Kastovsky and Aleksander Szwedek). Berlin, New York City, and Amsterdam. pp. 183–192.

Austerlitz, Robert (1990): Typology in the Service of Internal Reconstruction: Saxalin Nivx. In: *Language Typology 1987. Systematic Balance in Language. Papers from the Linguistic Typology Symposium*, [Berkeley 1987] (ed. by Winfred Philipp Lehmann). Amsterdam and Philadelphia. pp. 17–34.

Austerlitz, Robert (1994): Gilyak Internal Reconstruction, 3: Ligneous Matter. In: *Non-Slavic Languages of the USSR. Papers from the Fourth Conference on the Non-Slavic Languages of the USSR,* [Chicago 1985] (ed. by Howard Isaac Aronson and Bill Darden). Columbus/Ohio. pp. 229–233.

Bailey, Charles-James Nice (1969): The Integration of Linguistic Theory: Internal Reconstruction and the Comparative Method in Descriptive Linguistics. In: *Working Papers in Linguistics (University of Hawaii)* 2. pp. 85–121.

Baldi, Philip (1983): *An Introduction to the Indo-European Languages*. Carbondale/Illinois.

Bàrtoli, Matteo Giulio (1925): *Introduzione alla neolinguistica: prinzipi, scopi, metodi*. Geneva.

Bàrtoli, Matteo Giulio (1935): Analogie di metodo fra la storia dei linguaggi e delle tradizioni popolari. In: *Atti del III Congresso internazionale dei linguisti,* [Rome 1933] (ed. by Bruno Migliorini and Vittore Pisani). Florence. pp. 419–426.

Bauer, Brigitte (2009): Residues as an Aid in Internal Reconstruction. In: *Internal Reconstruction in Indo-European. Methods, Results, and Problems. Section Papers from the XVI International Conference on Historical Linguistics,* [Copenhagen 2003] (ed. by Jens Elmegård Rasmussen and Thomas Olander). Copenhagen. pp. 17–31.

Baxter, William Hubbard & Ramer, Alexis Manaster (2000): Beyond Lumping and Splitting: Probabilistic Issues in Historical Linguistics. In: *Time Depth in Historical Linguistics* 1. pp. 167–188.

Bech, Gunnar (1963): *Die Entstehung des schwachen Präteritums* (= Historisk-filosofiske Meddelelser udgivet af Det Kongelige Danske Videnskabernes Selskab 40/4). Copenhagen.

Beekes, Robert (1989): The Nature of Proto-Indo-European Laryngeals. In: The New Sound of Indo-European: Essays in Phonological Reconstruction (ed. by Theo Vennemann genannt Nierfeld). Berlin. pp. 23–33.

Beekes, Robert (1990): *Vergelijkende taalwetenschap. Een inleiding in de vergelijkende Indo-europese taalwetenschap*. Utrecht.

Beel, Joeran / Gipp, Bela / Langer, Stefan & Breitinger, Corinna (2016): Paper Recommender Systems: A Literature Survey. In: *International Journal on Digital Libraries* 17(4). pp. 305–338.

Benveniste, Émile (1935): *Origines de la formation des noms en indo-européen*. Paris.

Birnbaum, Henrik (1970): *Problems of Typological and Genetic Linguistics Viewed in a Generative Framework* (= Janua Linguarum Series Minor 106). The Hague and Paris.

Birnbaum, Henrik (1977): *Linguistic Reconstruction: Its Potential and Limitations in New Perspective* (= The Journal of Indo-European Studies / Monograph 2). Washington, D.C.

Blažek, Václav & Bičanová, Lenka (2014): Indo-European Personal Pronouns: Limits of Their Internal Reconstruction. In: *Linguistica Brunensia* 62(1). pp. 29–55.

Blevins, Juliette (2018): *Advances in Proto-Basque Reconstruction with Evidence for Proto-Indo-European-Euskarian Hypothesis* (= Routledge Studies in Historical Linguistics 2). New York City.

Bloomfield, Leonard (1925): On the Sound System of Central Algonkian. In: *Language* 1(4). pp. 130–156.

Bloomfield, Leonard (1933): *Language*. New York City.

Bloomfield, Leonard (1939): Menomini Morphophonemics. In: *Études phonologiques dédiées à la mémoire de M. le Prince N.S. Trubetzkoy* (= Travaux du Cercle Linguistique de Prague 8). Prague.

Bonfante, Giuliano (1945): On Reconstruction and the Linguistic Method. In: *Word* 1(1–2). pp. 83–91 (part 1) and pp. 132–161 (part 2).

Boretzky, Norbert (1975): Laryngaltheorie und innere Rekonstruktion. In: *Indogermanische Forschungen* 80. pp. 47–61.

Boretzky, Norbert (1977): *Einführung in die historische Linguistik* (= Rororo-Studium 108). Hamburg.

Borgström, Carl Hjalmar (1954): Internal Reconstruction of Pre-Indo-European Word-Forms. In: *Word* 10(2–3). pp. 275–287.

Bouchard-Côté, Alexandre / Hall, David / Griffiths, Thomas & Klein, Dan (2013): Automated Reconstruction of Ancient Languages Using Probabilistic Models of Sound Change. In: *Proceedings of the National Academy of Science of the United States of America* 111(11). pp. 4224–4229.

Bouckaert, Remco / Lemey, Philippe / Dunn, Michael / Greenhill, Simon / Alekseyenko, Alexander Vladimirovich / Drummond, Alexei / Gray, Russell

David / Suchard, Marc Adam & Atkinson, Quentin Douglas (2012): Mapping the Origins and Expansion of the Indo-European Language Family. In: *Science* 337(6097). pp. 957–960.

Brants, Sabine / Dipper, Stefanie / Eisenberg, Peter / Hansen, Silvia / König, Esther / Lezius, Wolfgang / Rohrer, Christian / Smith, George & Uszkoreit, Hans (2004): TIGER: Linguistic Interpretation of a German Corpus. In: *Journal of Language and Computation* 2(4). pp. 597–620.

Braune, Wilhelm & Heidermanns, Frank (2004): *Gotische Grammatik. Mit Lesestücken und Wörterverzeichnis* (= Sammlung kurzer Grammatiken germanischer Dialekte A1). Fully revised 20th edition. Tübingen.

Braune, Wilhelm & Reiffenstein, Ingo (2004): *Althochdeutsche Grammatik I. Laut- und Formenlehre* (= Sammlung kurzer Grammatiken germanischer Dialekte A 5/1). Fully revised 15th edition. Tübingen.

Bremer, Otto (1894): Relative Sprachchronologie. In: *Indogermanische Forschungen* 4. pp. 8–31.

Brown, Cecil / Holman, Eric / Wichmann, Søren / Velupillai, Viveka & Cysouw, Michael (2008): Automated Classification of the World's Languages. In: *Sprachtypologie und Universalienforschung* 61(4). pp. 285–308.

Brown, Cecil / Holman, Eric & Wichmann, Søren (2013): Sound Correspondences in the World's Languages. In: *Language* 89(1). pp. 4–29.

Bullinaria, John Andrew & Levy, Joseph (2007): Extracting Semantic Representations from Word Co-Occurrence Statistics: A Computational Study. In: *Behavior Research Methods* 39(3). pp. 510–526.

Burnage, Gavin (1995): CELEX: A Guide for Users. In: *The CELEX Lexical. Database, Release 2 (CD-ROM)* (ed. by Rolf Harald Baayen, Richard Piepenbrock & Léon Gulikers [online version, page http://www.kun.nl/celex/]). Philadelphia.

Bynon, Theodora (1977): *Historical Linguistics*. Cambridge/UK.

Campbell, Lyle (2013): *Historical Linguistics*. 3rd edition. Edinburgh.

Campbell, Lyle & Grondona, Verónica (2007): Internal Reconstruction in Chulupí (Nivaclé). In: *Diachronica* 24(1). pp. 1–29.

Chafe, Wallace (1959): Internal Reconstruction in Seneca. In: *Language* 35(3). pp. 477–495.

Carstensen, Kai-Uwe / Ebert, Christian / Ebert, Cornelia / Jekat, Susanne Johanna / Langer, Hagen & Klabunde, Ralf (2010): *Computerlinguistik und Sprachtechnologie: Eine Einführung*. 3rd edition. Heidelberg.

Chen, Matthew (1976): Relative Chronology: Three Methods of Reconstruction. In: *Journal of Linguistics* 12(2). pp. 209–258.

Chen, Matthew & Wang, William Shi-Yuan (1975): Sound Change: Actuation and Implementation. In: *Language* 51. pp. 255–281.

Chomsky, Noam & Halle, Morris (1968): The Sound Pattern of English. New York City.

Church, Kenneth Ward & Hanks, Patrick (1990): Word Association Norms, Mutual Information, and Lexicography. In: *Computational linguistics* 16(1). pp. 22–29.

Chvátal, Vacláv & Sankoff, David (1975): Longest Common Subsequences of Two Random Sequences. In: *Journal of Applied Probability* 12(2). pp. 306–315.

Comrie, Bernard (1980): Morphology and Word Order Reconstruction: Problems and Prospects. In: *Historical Morphology* (ed. by Jacek Fisiak). The Hague, Paris, New York City. pp. 83–96.

Comrie, Bernard (2003): Reconstruction, Typology and Reality. In: *Motives for Language Change* (ed. by Raymond Hickey). Cambridge/UK. pp. 243–257.

Cook, Eung-Do (1974): Internal Evidence for the Evolution of Number Categories in Sarcee. In: *Historical Linguistics I. Syntax, Morphology, Internal and Comparative Reconstruction. Proceedings of the First International Conference on Historical Linguistics*, [Edinburgh 1973] (ed. by John Mathieson Anderson and Charles Jones). Amsterdam. pp. 101–115.

Creutz, Mathias & Lagus, Krista (2005): Unsupervised Morpheme Segmentation and Morphology Induction from Text Corpora Using Morfessor 1.0. In: *Computer and Information Science, Report A81, Helsinki University of Technology*. pp. 1–27.

Croft, William / Bhattacharya, Tanmoy / Kleinschmidt, Dave / Smith, Eric & Jaeger, Florian (2011): Greenbergian Universals, Diachrony, and Statistical Analyses. In: *Linguistic Typology* 15(2). pp. 433–453.

Crowley, Terry & Bowern, Claire (2010): *An Introduction to Historical Linguistics*. 4th edition. Oxford and New York City.

Curtius, Georg (1873): *Das Verbum der griechischen Sprache seinem Bau nach dargestellt*, vol. 1. Leipzig.

de Boer, Bart (2000): Self-Organization in Vowel Systems. In: *Journal of Phonetics* 28(4). pp. 441–465.

de la Fuente, José Andrés Alonso (2017): On Internal Reconstruction, Morphological Restructuration and Nominal Classes in Manchu. In: *Acta Orientalia Academiae Scientiarum Hungaricae* 70(3). pp. 347–370.

Delbrück, Berthold (1880): *Einleitung in das Sprachstudium: ein Beitrag zur Geschichte und Methodik der vergleichenden Sprachforschung* (= Bibliothek indogermanischer Grammatiken 4). Leipzig.

Dellert, Johannes (2019): *Information-Theoretic Causal Inference of Lexical Flow* (= Language Variation 4). Berlin.

Dellert, Johannes / Daneyko, Thora / Münch, Alla / Ladygina, Alina / Buch, Armin / Clarius, Natalie / Grigorjew, Ilja / Balabel, Mohamed / Boga, Hizniye Isabella / Baysarova, Zalina / Mühlenbernd, Roland / Wahle, Johannes & Gerhard Jäger (2020): NorthEuraLex: A Wide-Coverage Lexical Database of Northern Eurasia. In: *Lang Resources & Evaluation* 54(1). pp. 273–301.

Derbolav, Josef (1972): *Platons Sprachphilosophie im Kratylos und in den späteren Schriften* (= Impulse der Forschung 10). Darmstadt.

Derwing, Bruce (1973): *Transformational Grammar as a Theory of Language Acquisition. A Study in the Empirical, Conceptual and Methodological Foundations of Contemporary Linguistics* (= Cambridge Studies in Linguistics 10). Cambridge/UK.

Dezső, László (1980): Grammatical Typology and Protolanguages. In: *Linguistic Reconstruction and Indo-European Syntax: Proceedings of the 'Indogermanische Gesellschaft'*, [Pavia 1979] (ed. by Paolo Ramat). Amsterdam. pp. 17–26.

Dixon, Robert Malcolm Ward (1994): *Ergativity* (Cambridge Studies in Linguistics 69). Cambridge/UK.

Dixon, Robert Malcolm Ward (2001): Internal Reconstruction of Tense-Modal Suffixes in Jarawara. In: *Diachronica* 18(1). pp. 3–30.

Dolgopolsky, Aharon Borissovich (1964): Гипотеза древнейшего родства языковых семей северной Евразии с вероятностной точки зрения [Gipoteza drevneyshego rodstva yazykovyx semey severnoy Evrazii s veroyatnostnoy tochki zreniya]. In: *Вопросы Языкознания* 2 [*Voprosy YAzykoznaniya* 2]. pp. 53–63.

Dolgopolsky, Aharon Borissovich (1989): Cultural Contacts of Proto-Indo-European and Proto-Indo-Iranian with Neighbouring Languages. In: *Folia Linguistica Historica* 8 (1–2). pp. 3–36.

Dunkel, George Eugene (2014): *Lexikon der indogermanischen Partikel und Pronominalstämme*, vol. 2: *Lexikon*. Heidelberg.

Dunn, Michael (2012): *Indo-European Lexical Cognacy Database (IELex)* [page http://ielex.mpi.nl/]. Nijmegen.

Eichner, Heiner (1988): Sprachwandel und Rekonstruktion. In: *Akten der 13. Österreichischen Linguistentagung,* [Graz 1985] (ed. by Christian Zinko). Graz. pp. 10–40.

Eisenberg, Peter (2011): *Das Fremdwort im Deutschen*. Berlin and New York City.

Euler, Wolfram & Badenheuer, Konrad (2009): *Sprache und Herkunft der Germanen. Abriss des Protogermanischen vor der Ersten Lautverschiebung*. Hamburg and London.

Ewen, Colin & Hulst, Harry van der (2001): *The Phonological Structure of Words. An Introduction* (= Cambridge Textbooks in Linguistics). Cambridge/UK.

Ferguson, Charles Albert (1966): Assumptions about Nasals: A Sample Study in Phonological Universals. In: *Universals of Language* (ed. by Joseph Harold Greenberg). 2nd edition. Cambridge/Massachusetts and London.

Fortson, Benjamin Wynn (2010): *Indo-European Language and Culture: An Introduction*. 2nd edition (= Blackwell textbooks in linguistics 19). Chichester/West Sussex.

Fowkes, Robert Allen (1950): Synchronic Method and Welsh Consonantism. In: *Word* 9. pp. 142–145.

Fox, Anthony (1995): *Linguistic Reconstruction: An Introduction to Theory and Method*. Oxford.

Freeze, Ray (1976): Internal Reconstruction in Hopi Phonology. In: *Current Progress in Historical Linguistics. Proceedings of the Second International Conference on Historical Linguistics*, [Tucson 1976] (ed. by William Morrison Christie, Jr.). Amsterdam. pp. 195–202.

Fronzaroli, Pelio (1970): "Ricostruzione interna" del verbo semitico in alcuni studi recenti. In: *Mille. I Dibattiti del circolo linguistico fiorentino 1945–1970* (= Accademia Toscana di Scienze e Lettere "La Colombaria" 16). Florence. pp. 71–85.

Garrett, Andrew (2015): Sound Change. In: *The Routledge Handbook of Historical Linguistics* (ed. by Claire Bowern and Bethwyn Evans). Abingdon-on-Thames and New York City. pp. 227–248.

Geisler, Hans & List, Johann-Mattis (2009): *Beautiful Trees on Unstable Ground: Notes on the Data Problem in Lexicostatistics*. [presentation held at the Arbeitstagung der Indogermanischen Gesellschaft]. Würzburg.

Givón, Talmy (1971): Historical Syntax and Synchronic Morphology. An Archaeologist's Field-Trip. In: *Papers from the Seventh Regional Meeting, Chicago Linguistic Societey*, [Chicago 1971]. Chicago. pp. 394–415.

Givón, Talmy (1990): *Syntax: A Functional-Typological Introduction*, vol. 2. Amsterdam and Philadelphia.

Givón, Talmy (1999a): Internal Reconstruction: As Method, as Theory. In: *Typological Studies in Language* 43. pp. 107–159.

Givón, Talmy (1999b): Internal Reconstruction, on Its Own. In: *Language Change and Typological Variation. In Honor of Winfried P. Lehmann on the Occasion of His 83rd Birthday* (ed. by Edgar Charles Polomé and Carol Faith Justus). Washington, D.C. pp. 86–129.

Gmür, Remo (1986): *Das Schicksal von F. de Saussure „Mémoire": eine Rezeptionsgeschichte.* (= Arbeitspapier der Universität Bern, Institut für Sprachwissenschaft 21). Bern.

Grammont, Maurice (1901): Onomatopées et mots expressifs. In: *Revue des Langues Romanes* 44. pp. 97–159.

Graßmann, Hermann (1860): Ueber die verbindung der stummen consonanten mit folgendem v und die davon abhängigen erscheinungen. In: *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen* 9. pp. 1–35.

Gray, Russell David & Atkinson, Quentin Douglas (2003): Language-Tree Divergence Times Support the Anatolian Theory of Indo-European Origin. In: *Nature* 426(6965). pp. 435–439.

Greenberg, Joseph Harold (1966): Synchronic and Diachronic Universals in Phonology. In: *Language* 42(2). pp. 508–517.

Greenberg, Joseph Harold (1978): Diachrony, Synchrony and Language Universals. In: *Universals of Human Language, vol. 1: Method and Theory* (ed. by Joseph Harold Greenberg, Charles Albert Ferguson, and Edith Andrea Moravcsik). Stanford. pp. 61–91.

Greenhill, Simon (2015): TransNewGuinea.org: An Online Database of New Guinea Languages. In: *PLoS ONE* 10. pp. 1–17.

Greenhill, Simon / Blust, Robert & Gray, Russell David (2008): The Austronesian Basic Vocabulary Database: From Bioinformatics to Lexiomics. In: *Evolutionary Bioinformatics* 4. pp. 271–283.

Greenhill, Simon / Currie, Thomas & Gray, Russell David (2009): Does Horizontal Transmission Invalidate Cultural Phylogenies? In: *Proceedings of the Royal Society B: Biological Sciences* 276(1665). pp. 2299–2306.

Haas, Mary Rosamond (1969): Internal Reconstruction of the Nootka-Nitinat Pronominal Suffixes. In: *International Journal of American Linguistics* 35(2). pp. 108–124.

Haas, Mary Rosamond (1980): Notes on Karok Internal Reconstruction. In: *American Indian and Indoeuropean Studies: Papers in Honor of Madison S. Beeler* (ed. by Kathryn Klar, Margaret Langdon, and Shirley Silver). The Hague and New York City. pp. 67–76.

Hajnal, Ivo (2016): Induktive versus abduktive Rekonstruktion. In: *Indogermanische Forschungen* 121(1). pp. 435–454.

Hall, Robert Anderson, Jr. (1950): The Reconstruction of Proto-Romance. In: *Language* 26(1). pp. 6–27.

Halle, Morris (1961): On the Role of Simplicity in Linguistic Descriptions. In: *Structure of Language and Its Mathematical Aspects. Proceedings of Symposia in Applied Mathematics*, [New York City 1960] (ed. by Roman Jakobson). Providence/Rhode Island. pp. 89–94.

Halle, Morris (1964): On the Bases of Phonology. In: *The Structure of Language: Readings in the Philosophy of Language* (ed. by Jerry Alan Fodor and Jerrold Jacob Katz). Englewood Cliffs/New Jersey. pp. 324–333.

Hamp, Birgit & Feldweg, Helmut (1997): GermaNet – A Lexical-Semantic Net for German. In: *Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications. ACL/EACL-97 Workshop Proceedings,* [Madrid 1997] (ed. by Piek Vossen, Geert Adriaens, Nicoletta Calzolari, Antonio Sanfilippo, and Yorick Wilks). Somerset/New Jersey. pp. 9–15.

Harris, Martin (1995): Word Order. In: *Syntax. An International Handbook of Contemporary Research*, vol. 2 (ed. by Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, and Theo Vennemann genannt Nierfeld). Berlin and New York City. pp. 1216–1226.

Hartigan, John Anthony & Wong, Manchek (1979): Algorithm AS 136: A K-Means Clustering Algorithm. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 28(1). pp. 100–108.

Hasil, Jiří (2008): Die Stratifikation des Tschechischen als Zertifizierungsproblem. In: *Europaeische Sprachpolitik und Zertifizierung des Polnischen und Tschechischen* (ed. by Danuta Rytel-Kuc and Jolanta Tambor). Frankfurt/Main. pp. 134–139.

Haspelmath, Martin (2001): The European Linguistic Area: Standard Average European. In: *Language Typology and Language Universals* (ed. by Martin Haspelmath, Wulf Oesterreicher, and Wolfgang Raible). Berlin. pp. 1492–1510.

Haßler, Gerda & Neis, Cordula (2009): *Lexikon sprachtheoretischer Grundbegriffe des 17. und 18. Jahrhunderts*. Berlin.

Heitsch, Ernst (1984): *Willkür und Problembewusstsein in Platons Kratylos* (= Abhandlungen der Geistes- und Sozialwissenschaften Klasse / Akademie der Wissenschaften und der Literatur Mainz 11). Stuttgart.

Henrich, Verena & Hinrichs, Erhard (2010): GernEdiT – The GermaNet Editing Tool. In: *Proceedings of the Seventh Conference on International Language Resources and Evaluation*, [Valletta 2010] (ed. by Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Jan Odijk, Stellos Piperidis, Joseph Mariani, Mike Rosner, and Daniel Tapias). Paris. pp. 2228–2235.

Henry, Victor (1896): Revue du 27. Janvier. Schmidt, J.: Nouvelles Recherches linguistiques (Kritik der Sonantentheorie, 1895). In: *Revue Critique d'Histoire et de Littérature* 41. pp. 58–60.

Hermann, Eduard (1907): Über das Rekonstruieren. In: *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 41. pp. 1–64.

Hewson, John (1974): Comparative Reconstruction on the Computer. In: *Historical Linguistics I. Syntax, Morphology, Internal and Comparative Reconstruction. Proceedings of the First International Conference on Historical Linguistics*, [Edinburgh 1973] (ed. by John Mathieson Anderson and Charles Jones). Amsterdam. pp. 191–197.

Hewson, John (1989): Computer Analysis of Proto-Algonquian Word Formatives. In: *Algonquian Papers-Archive* 20. pp. 154-160.

Hewson, John (1993): *A Computer-Generated: Dictionary of Proto-Algonquian* (= National Museum of Civilization, Mercury Series, Canadian Ethnology Service Paper 125). Hull/Quebec.

Hill, Nathan Wayne & List, Johann-Mattis (2017): Challenges of Annotation and Analysis in Computer-Assisted Language Comparison: A Case Study on Burmish Languages. In: *Yearbook of the Poznań Linguistic Meeting* 3(1). pp. 47–76.

Hiller, Marion (2001): *Das „zwitterhafte" Wesen des Wortes. Eine Interpretation von Platons Dialog „Kratylos"* (= Phainomena 12). Tübingen.

Hirschfeld, Ursula & Stock, Eberhard (2014): Wie kommt die Aussprache ins (Aussprache-) Wörterbuch? Methoden, Probleme und Ergebnisse normphonetischer Untersuchungen zur deutschen Standardaussprache. In: *Lexicographica* 30(1). pp. 262–290.

Hjelmslev, Louis (1966): *Le Langage. Une introduction. Préface de Algirdas Julien Greimas* (= Arguments 28). Translation of *Sproget. En introduktion*, Berlin and Copenhagen 1963 (translated by Michel Olsen). Paris.

Hock, Hans Henrich (1986): *Principles of Historical Linguistics* (= Trends in Linguistics. Studies and Monographs 34). Berlin, New York City, and Amsterdam.

Hock, Hans Henrich & Joseph, Brian Daniel (2019): *Language History, Language Change, and Language Relationship. An Introduction to Historical and Comparative Linguistics.* Revised 3rd edition. Berlin and Boston.

Hock, Wolfgang (2013): Interne Rekonstruktion und Akzentparadigma. In: *Indo-European Accent and Ablaut* (ed. by Götz Keydana, Paul Widmer, and Thomas Olander). Copenhagen. pp. 1–30.

Hockett, Charles Francis (1948): Implications of Bloomfield's Algonquian Studies. In: *Language* 24(1). pp. 117–131.

Hockett, Charles Francis (1958): *A Course in Modern Linguistics*. New York City.

Hoenigswald, Henry Max (1944): Internal Reconstruction. In: *Studies in Linguistics* 2(4). pp. 78–87.

Hoenigswald, Henry Max (1946): Sound Change and Linguistic Reconstruction. In: *Language* 22. pp. 138–143.

Hoenigswald, Henry Max (1950): The Principal Step in Comparative Grammar. In: *Language* 26. pp. 357–364.

Hoenigswald, Henry Max (1965): Language Change and Linguistic Reconstruction. Chicago and London.

Hoenigswald, Henry Max (1974): Internal Reconstruction and Context. In: *Historical Linguistics II. Theory and Description in Phonology. Proceedings of the First International Conference on Historical Linguistics*, [Edinburgh 1973] (ed. by John Mathieson Anderson and Charles Jones). Amsterdam. pp. 189–201.

Hoenigswald, Henry Max (1992): Comparative Method, Internal Reconstruction, Typology. In: *Reconstructing Languages and Cultures* (ed. by Edgar Charles Polomé and Werner Winter). Berlin and New York City. pp. 23–34.

Hoijer, Harry (1969): Internal Reconstruction in Navaho. In: *Word* 25. pp. 155–159.

Hruschka, Daniel Jacob / Bradford, Simon / Smith, Eric / Wilkins, Jon / Meade, Andrew / Pagel, Mark & Bhattacharya, Tanmoy (2015): Detecting Regular Sound Change in Linguistics as Events of Concerted Evolution. In: *Current Biology* 25(1). pp. 1–9.

Humboldt, Friedrich Wilhelm von (1822): Ueber das Entstehen der grammatischen Formen und ihren Einfluss auf die Ideenentwicklung. In: *Die sprachphilosophischen Werke Wilhelms von Humboldt* (ed. by Heymann Steinthal). Berlin [reprint Berlin 1884]. pp. 67–101.

Hyllested, Adam (2009): Internal Reconstruction vs. External Comparison: The Case of the Indo-Uralic Larnygeals. In: *Internal Reconstruction in Indo-European: Methods, Results, and Problems. Section Papers from the XVI International Conference on Historical Linguistics,* [Copenhagen 2003] (ed. by Jens Elmegård Rasmussen and Thomas Olander). Copenhagen. pp. 111–136.

Igartua, Iván (2015): From Cumulative to Separative Exponence in Inflection: Reversing the Morphological Cycle. In: *Language* 91(3). pp. 676–722.

Isebaert, Lambert (1991): The Method of Internal Reconstruction in Linguistics: An Additional Bibliography. In: *Folia Linguistica Historica* 12(1–2). pp. 219–224.

Ivanov, Vlačeslav Vsevolodovič & Toporov, Vladimir Nikolayevič (1961): К постановке вопроса о древнейших отношениях балтийских и славянских языков [K postanovke voprosa o drevnejshix otnoshenijax baltijskix i slavjanskix jazykov] In: *Исследования по славянскому языкознанию* [*Issledovanija po slavjanskomu jazykoznaniju]* (ed. by Nikita Il'ič TOLSTOJ). Moscow. pp. 273–305.

Jacques, Guillaume (2017): A Reconstruction of Proto-Kiranti Verb Roots. In: *Folia Linguistica* 51. pp. 177–215.

Jäger, Gerhard (2015): Support for Linguistic Macrofamilies from Weighted Sequence Alignment. In: *Proceedings of the National Academy of Sciences* 112(41). pp. 12752–12757.

Jäger, Gerhard (2019): Computational Historical Linguistics. In: *Theoretical Linguistics* 45(3–4). pp. 151–182.

Jäger, Gerhard & List, Johann-Mattis (2018): Using Ancestral State Reconstruction Methods for Onomasiological Reconstruction in Multilingual Word Lists. In: *Language Dynamics and Change* 8(1). pp. 22–54.

Jakobson, Roman (1941): *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala.

Jakobson, Roman (1958): Typological Studies and Their Contribution to Historical Comparative Linguistics. In: *Proceedings of the 8th International Congress of Linguistics,* [Oslo 1957] (ed. by Eva Sivertsen). Oslo. pp. 17–35.

Jakobson, Roman (1975): Three Linguistic Abstracts. In: *Studii şi cercetări lingvistice* 26. pp. 505–507.

Janda, Richard David & Joseph, Brian Daniel (2003): On Language, Change, and Language Change – Or, Of History, Linguistics, and Historical Lingustics. In: *The Handbook of Historical Linguistics* (ed. by Brian Daniel Joseph and Richard David Janda). Malden. pp. 3–180.

Jeffers, Robert John (1976): Syntactic Change and Reconstruction. In: *Current Progress in Historical Linguistics. Proceedings of the 2nd International Conference on Historical Linguistics,* [Tucson 1976]. (ed. by William Morrison Christie, Jr.). Amsterdam, New York City, and Oxford. pp. 1–16.

Jeffers, Robert John (1977): Morphological Reanalysis and Analogy: Two Case Histories from Latin and Greek. In: *Lingua* 41. pp. 13–24.

Jeffers, Robert John & Lehiste, Ilse (1979): Internal Reconstruction. In: *Principles and Methods for Historical Linguistics*. Cambridge/Massachusetts. pp. 37–54.

Jespersen, Otto (1922): *Language: Its Nature, Development, and Origin*. London.

Johansson, Karl Ferdinand (1900): Anlautendes indogerman. b. In: *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen*, 36(3). pp. 342–390.

Jordan, Fiona / Gray, Russel David / Greenhill, Simon & Mace, Ruth (2009): Matrilocal Residence Is Ancestral in Austronesian Societies. In: *Proceedings of the Royal Society B* 276(1664). pp. 1957–1964.

Joseph, Brian Daniel (2010): Internal Reconstruction. In: *Bloomsbury Companion to Historical Linguistics* (ed. by Silvia Luraghi and Vit Bubenik). London and New York City. pp. 52–58.

Kapović, Mate (2017): Proto-Indo-European Morphology. In: *The Indo-European Languages* (ed. by Mate Kapović). 2nd edition. Abingdon-on-Thames and New York City. pp. 61–110.

Kassian, Alexey / Zhivlov, Mikhail & Starostin, George (2015): Proto-Indo-European-Uralic comparison from the probabilistic point of view. In: *Journal of Indo-European Studies* 43(3–4). pp. 301–347.

Katičić, Radoslav (1966): Der Entsprechungsbegriff in der vergleichenden Laut- und Formenlehre. In: *Indogermanische Forschungen* 71. pp. 203–220.

Kay, Martin (1964): *The Logic of Cognate Recognition in Historical Linguistics* (= Research memorandum / Rand Corporation 4224). Santa Monica.

Keller, Rudi (1994): *On Language Change. The Invisible Hand in Language*. Translation of *Sprachwandel. Von der unsichtbaren Hand in der Sprache*, Tübingen 1990 (translated by Brigitte Nerlich). London and New York City.

Kelly, David Hilary (1960): Internal Reconstruction of Juncture for Proto-Greek. In: *Language* 36. pp. 351–353.

King, Robert Desmond (1971): *Historische Linguistik und generative Grammatik* (= Schwerpunkte Linguistik und Kommunikationswissenschaft 5). Translation of *Historical Linguistics and Generative Grammar*, Englewood Cliffs/New Jersey 1969 (translated and edited by Steffen Stelzer). Frankfurt/Main.

Kiparsky, Paul (1968): Linguistic Universals and Linguistic Change. In: *Universals in Linguistic Theory* (ed. by Emmon Bach and Robert Thomas Harms). New York City. pp. 170–202.

Kiparsky, Paul (2014): New Perspectives in Historical Linguistics. In: *The Routledge Handbook of Historical Linguistics* (ed. by Claire Bowern and Bethwyn Evans). London. pp. 64–102.

Kondrak, Grzegorz (2002): Determining Recurrent Sound Correspondences by Inducing Translation Models. In: *Proceedings of the 19th International Conference on Computational Linguistics*, vol. 1, [Taipei 2002] (ed. by Shu-Chuan Tseng, Tsuei-Er Chen, and Yi-Fen Liu). San Francisco. pp. 1–7.

Kondrak, Grzegorz & Sherif, Tarek (2006): Evaluation of Several Phonetic Similarity Algorithms on the Task of Cognate Identification. In: *Proceedings of the Workshop on Linguistic Distances*, [Sydney 2006] (ed. by John Nerbonne and Erhard Hinrichs). Stroudsburg/Pennsylvania. pp. 43–50.

Koskenniemi, Kimmo (2013): Finite-State Relations between Two Historically Closely Related Languages. In: *Proceedings of the Workshop on Computational Historical Linguistics at NODALIDA 2013*, [Oslo 2013] (ed. by Þórhallur Eyþórsson, Lars Borin, Dag Haug, and Eiríkur Rögnvaldsson). Linköping. pp. 43–53.

Kümmel, Martin Joachim (2007): *Konsonantenwandel. Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden.

Kurylowicz, Jerzy (1964): On the Methods of Internal Reconstruction. In: *Proceedings of the 9th International Congress of Linguists*, [Cambridge/Massachusetts 1962] (ed. by Horace Gray Lunt). The Hague. pp. 9–36.

Kurylowicz, Jerzy (1973): Internal Reconstruction. *In: Current Trends in Linguistics,* vol 11*: Diachronic, Areal, and Typological Linguistics* (ed. by Thomas Albert Sebeok). pp. 63–92. The Hague and Paris.

Labov, William (1981): Resolving the Neogrammarian Controversy. In: *Language* 57(2). pp. 267–308.

Landerman, Peter (1997): Internal Reconstruction in Aymara and Quechua. In: *The Life of Language. Papers in Linguistics in Honor of William Bright* (ed. by Jane Hassler Hill, Purushottam Jivanji Mistry, and Lyle Campbell). Berlin and New York City. pp. 35–58.

Langdon, Margaret (1997): Notes on Highland Chontal Internal Reconstruction. In: *The Hokan, Penutian & Harrington Conferences and the Mary R. Haas Memorial*, [Berkeley 1996] (ed. by Leanne Hinton). Berkeley. pp. 25–34.

Lass, Roger (1975): Internal Reconstruction and Generative Phonology. In: *Transactions of the Philological Society* 74(1). pp. 1–26.

Lass, Roger (1980): *On Explaining Language Change* (= Cambridge Studies in Linguistics 27). Cambridge/UK, London, New York City, and Melbourne.

Lass, Roger (1997): *Historical Linguistics and Language Change* (= Cambridge Studies in Linguistics 81). Cambridge/UK, New York City, and Melbourne.

Lass, Roger & Anderson, John Mathieson (1975): Old English Phonology (= Cambridge Studies in Linguistics 14). Cambridge/UK, London, New York City, and Melbourne.

Latta, Fletcher Christian (1978): *Methods for Internal Reconstruction* (Doctoral dissertation, The Ohio State University). Columbus/Ohio.

Lehmann, Winfred Philipp (1961): Выводы о протоиндоевропейской глагольной системе, основанные на внутреннем анализе санскрита [Vyvody o protoindoyevropeyskoy glagol'noy sisteme, osnovannyye na vnutrennem analize sanskrita]. In: *Вопросы Языку* 2 [*Vosprosy YAzyku* 2]. pp. 24–27.

Lehmann, Winfred Philipp (1962): Historical Linguistics: An Introduction. New York City.

Lehmann, Winfred Philipp (1989): Problems in Proto-Indo-European Grammar: Residues from Pre-Indo-European Active Structure. In: *General Linguistics* 29. pp. 228–246.

Levenshtein, Vladimier Iosifovich (1966): Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. In: *Soviet Physics-Doklady* 10(8). pp. 707–710.

Lidstone, George James (1920): Note on the General Case of the Bayes-Laplace Formula for Inductive or a Posteriori Probabilities. In: *Transactions of the Faculty of Actuaries* 8. pp. 182–192.

Lightfoot, David William (1979): *Principles of Diachronic Syntax* (= Cambridge Studies in Linguistics 23). Cambridge/UK, London, New York City, and Melbourne.

Linell, Per (1974): *Problems of Psychological Reality in Generative Phonology: A Critical Assessment* (= Reports from Uppsala University Department of Linguistics 4). Uppsala.

List, Johann-Mattis (2010): SCA: Phonetic Alignment Based on Sound Classes. In: *New Directions in Logic, Language and Computation* (ed. by Marija Slavkovik and Daniel Lassiter). Berlin and Heidelberg. pp. 32–51.

List, Johann-Mattis (2012): Lexstat: Automatic Detection of Cognates in Multilingual Wordlists. In: *Proceedings of the EACL 2012 Joint Workshop of LINGVIS and UNCLH: Visualization of Linguistic Patterns and Uncovering Language History from Multilingual Resources*, [Avignon 2012] (ed. by MIRIAM BUTT, Sheelagh Carpendale, Gerald Penn, Jelena Prokić, and Michael Cysouw). Stroudsburg/Pennsylvania. pp. 117–125.

List, Johann-Mattis (2014): Investigating the Impact of Sample Size on Cognate Detection. In: *Journal of Language Relationship* 11(1). pp. 91–101.

List, Johann-Mattis (2018): Horizontal and Vertical Language Comparison. In: *Annual Contributions of the Genealogical World of Phylogenetic Networks* 7(6). pp. 28–33.

List, Johann-Mattis (2019): *Open Problems in Computational Historical Linguistics* [presentation held at the 24th International Conference of Historical Linguistics]. Canberra.

List, Johann-Mattis / Lopez, Philippe & Bapteste, Eric (2016): Using Sequence Similarity Networks to Identify Partial Cognates in Multilingual Wordlists. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics,* vol. 2: *Short Papers* [Berlin, 2016] (ed. by Katrin Erk and Noah Smith). Stroudsburg/Pennsylvania. pp. 599–605.

List, Johann-Mattis / Greenhill, Simon & Gray, Russell David (2017): The Potential of Automatic Word Comparison for Historical Linguistics. In: *PLoS ONE* 12(1). pp. 1–18.

Liu, Bing (2007): *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. 2nd edition. Berlin and Heidelberg.

Lloyd, Stuart (1982): *Least Squares Quantization in PCM*. In: *IEEE Transactions on Information Theory* 28(2). pp. 129–137.

Lounsbury, Floyd Gleda (1953): *Oneida Verb Morphology* (= Yale University Publications in Anthropology 48). New Haven and London.

Lovitt, Andrew (2007): Correcting Confusion Matrices for Phone Recognizers. In: *Rapport de recherche de l'IDIAP* 07-03. pp. 1–13.

Lühr, Rosemarie (1988): *Expressivität und Lautgesetz im Germanischen* (= Monographien zur Sprachwissenschaft 15). Heidelberg.

MacKay, David John Cameron (2003): *Information Theory, Inference, and Learning Algorithms.* Cambridge/UK.

Maher, John Peter (1977): The Paradox of Creation and Tradition in Grammar. Sound Pattern of a Palimpsest. In: *Papers on Language Theory and History,* vol. 1*: Creation and Tradition in Language* (ed. by John Peter Maher). Amsterdam. pp. 1–34.

McMahon, April & McMahon, Robert (2005): *Language Classification by Numbers*. Oxford and New York City.

Mallory, James Patrick & Adams, Douglas Quentin (2006): *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World*. New York City.

Marchand, James Woodrow (1956): Internal Reconstruction of Phonemic Split. In: *Language* 32(2). pp. 245–253.

Martin, Samuel Elmo (1996): *Consonant Lenition in Korean and the Macro-Altaic Question* (= Center for Korean Studies Monographs 19). Honolulu.

Martinet, André (1955): *Économie des changements linguistiques*. Bern.

Martinez Areta, Mikel (2013): Demonstratives and Personal Pronouns. In: *Basque and Proto-Basque. Language Internal and Typological Approaches to Linguistic Reconstruction* (ed. by Mikel Martínez Areta). Frankfurt/Main. pp. 283–321.

Matzel, Klaus (1962): Anlautendes þl- und fl- im Gotischen. In: *Die Sprache* 8. pp. 220–237.

Mayerthaler, Willi (1974): *Eine Einführung in die generative Phonologie* (= Romanistische Arbeitshefte 11). Tübingen.

Mayrhofer, Manfred (1980): Über sprachliche Rekonstruktionsmethoden. In: *Anzeiger der philosophisch-historischen Klasse der Österreichischen Akademie der Wissenschaften* 117(25). pp. 357–366.

Mayrhofer, Manfred (1982): Über griechische Vokalprothese, Laryngaltheorie und externe Rekonstruktion. In: *Serta Indogermanica. Festschrift für Günter Neumann zum 60. Geburtstag* (ed. by Johann Tischler). Innsbruck. pp. 177–192.

Mehendale, Madhukar Anant (1960): Limitations of the Method of Internal Reconstruction. In: *Indian Linguistics* 21. pp. 101–103.

Mehendale, Madhukar Anant (1963): Internal and Comparative Reconstruction. In: *Indian Linguistics* 24. pp. 41–47.

Meier-Brügger, Michael (2010): *Indogermanische Sprachwissenschaft*, Revised 9th edition (with contributions by Matthias Fritz and Manfred Mayrhofer). Berlin and New York City.

Meillet, Antoine (1925): *La méthode comparative en linguistique historique* (= Instituttet for Sammenlignende Kulturforskning 2). Oslo [reprint Paris 1954].

Meillet, Antoine (1937): *Introduction à l'étude comparative des langues indo-européennes*. Revised 8th edition. Paris [reprint Forge Village/Massachusetts 1964].

Meiser, Gerhard (2003): *Veni Vidi Vici. Die Vorgeschichte des lateinischen Perfektsystems* (= Zetemata – Monographien zur Klassischen Altertumswissenschaft 113). München.

Meiser, Gerhard (2010): *Historische Laut- und Formenlehre der lateinischen Sprache*. 3rd edition. Darmstadt.

Meyer, Christian & Gurevych, Iryna (2012): OntoWiktionary: Constructing an Ontology from the Collaborative Online Dictionary Wiktionary. In: *Semi-Automatic Ontology Development: Processes and Resources* (ed. by Maria Teresa Pazienza and Armando Stellato). Hershey/Pennsylvania. pp. 131–161.

Michailovsky, Boyd (2012): Internal reconstruction of the Dumi Verb. Lexical Base and Stem Formation. In: *Linguistics of the Tibeto-Burman Area* 35(2). pp. 49–87.

Michelena, Luis (1977): *Fonética Histórica Vasca.* Revised 2nd edition. San Sebastián [reprint San Sebastian 1990].

Milewski, Tadeusz (1973): *Introduction to the Study of Language* (= Ianua Linguarum / Series Didactica 7). Translation of *Językoznawstwo*, Warsaw 1965 (translated by Maria Brochwicz). The Hague and Warsaw.

Miller, George Armitage (1995): WordNet: A Lexical Database for English. In: *Communications of the ACM* 38(11). pp. 39–41.

Miranda, Rocky Valerine (1975): Internal Reconstruction: Scope and Limits. In: *Lingua* 36(4). pp. 289–305.

Mitzka, Walther (1967): Mittelfränkische Konsonantenschwächung. In: *Zeitschrift für Mundartforschung* 34(3–4). pp. 254–257.

Møller, Hermann (1917): *Die semitisch-vorindogermanischen laryngalen Konsonanten* (= Det Kongelige Danske Videnskabernes Selskabs skrifter. Historisk og filosofisk Asdelning 7, vol. 4). Copenhagen.

Moran, Steven / Grossman, Eitan & Annemarie Verkerk (2020): Investigating Diachronic Trends in Phonological Inventories Using BDPROTO. In: *Language Resources and Evaluation* 55(1). pp. 1–25.

Naert, Pierre (1957): Sur la méthode de la « reconstruction interne ». In: *Studia Linguistica* 11(1-2). pp. 1–7.

Navigli, Roberto & Ponzetto, Simone Paolo (2010): BabelNet: Building a Very Large Multilingual Semantic Network. In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, [Uppsala 2010] (ed. by Jan Hajič, Sandra Carberry, Stephen Clark, and Joakim Nivre). Stroudsburg/Pennsylvania. pp. 216–225.

Needleman, Saul Ben & Wunsch, Christian (1970): A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. In: *Journal of Molecular Biology* 48(3). pp. 443–453.

Nelson, Michael (2020): *Statistics in Nutrition and Dietetics*. Hoboken/New Jersey and Chichester.

Nguyen, Kim Anh / Schulte im Walde, Sabine & Vu, Ngoc Thang (2016): Integrating Distributional Lexical Contrast into Word Embeddings for Antonym-Synonym Distinction. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics: proceedings of System Demonstrations*, [Berlin 2016] (ed. by Sameer Pradhan and Marianna Apidianaki). Stroudsburg/Pennsylvania. pp. 454–459.

Norton, Russel (2013): The Ama Dual Suffix: An Internal Reconstruction. In: *Nilo-Saharan – Models and Description* (ed. by Angelika Mietzner and Anne Storch). Cologne. pp. 113–122.

Oakes, Michael (2000): Computer Estimation of Vocabulary in a Protolanguage from Word Lists in Four Daughter Languages. In: *Journal of Quantitative Linguistics* 7(3). pp. 233–243.

O’Brien, Mary Grantham & Fagan, Sarah (2016): *German Phonetics and Phonology. Theory and Practice*. New Haven/Connecticut and London.

Obst, Wolfgang & Schleburg, Florian (2004): *Lehrbuch des Altenglischen*. Heidelberg.

Oertel, Hanns (1897): On the Character of Inferred Parent Languages. In: *The American Journal of Philology* 18(4). pp. 416–438.

Olsen, Birgit Anette (2010): Martinet’s Rule of Laryngeal Hardening: A Reappraisal. In: *Proceedings of the 21st Annual UCLA Indo-European Conference*, [Los Angeles 2009] (ed. by Stephanie Wroth Jamison, Harold Craig

Melchert, Brent Harmon Vine, and Angelo Mercado). Bremen. pp. 207–222.

Ono, Masataka / Miwa, Makoto & Sasaki, Yutaka (2015): Word Embedding-Based Antonym Detection Using Thesauri and Derivational Information. In: *Proceedings of the 2015 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, vol. 1: *Human Language Technologies*, [Denver 2015] (ed. by Rada Mihalcea, Joyce Yue Chai, and Anoop Sarkar). Denver. pp. 984–989.

Padó, Sebastian & Lapata, Mirella (2007): Dependency-Based Construction of Semantic Space Models. In: *Computational Linguistics* 33(2). pp. 161–199.

Pagel, Mark / Atkinson, Quentin Douglas Atkinson & Meade, Andrew (2007): Frequency of Word-Use Predicts Rates of Lexical Evolution Throughout Indo-European History. In: *Nature* 449(7163). pp. 717–720.

Pagliaro, Antonino (1930): *Sommario di linguistica arioeuropea*, vol. 1: *Cenni storici e questioni teoriche* (= Pubblicazioni della Scuola di Filologia Classica 2). Rome.

Paraschkewow, Boris (2004): *Wörter und Namen gleicher Herkunft und Struktur. Lexikon etymologischer Dubletten im Deutschen.* Berlin and New York City.

Paul, Hermann / Moser, Hugo & Schröbler, Ingeborg (1969): *Mittelhochdeutsche Grammatik* (= Sammlung kurzer Grammatiken germanischer Dialekte, series A, vol. 2). 20th edition. Tübingen.

Pedersen, Holger (1905): Die nasalpräsentia und der Slavische akzent. In: *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 38(3). pp. 297–421.

Peirce, Charles Sanders (1934): Lectures on Pragmatism. In: *Collected Papers of Charles Sanders Peirce*, vol. 5: *Pragmatism and pragmaticism* (ed. by Charles Hartshorne, Paul Weiss, and Arthur Walter Burks). Cambridge/Massachusetts. pp. 14–212.

Pereltsvaig, Asya & Lewis, Martin: *The Indo-European Controversy. Facts and Fallacies in Historical Linguistics*. Cambridge/UK.

Pisani, Vittore (1938): *Paleontologia linguistica*, vol. 1: *note critiche e metodologiche* (= Annali della Facoltà di Lettere e Filosofia dell'Università di Cagliari 9). Cagliari.

Pisani, Vittore (1940): *Geolinguistica e indoeuropeo: memoria* (= Memorie / Accademia nazionale dei Lincei. Classe di scienze morali, storiche e filologiche, series 6, vol. 9). Rome.

Pisani, Vittore (1974): Zum Problem labialer Wurzelerweiterungen, bzw. labialer Suffixe in den indogermanischen Sprachen. In: *Indogermanische Forschungen* 79. pp. 92–151.

Pokorny, Julius (1959): *Indogermanisches etymologisches Wörterbuch*. Bern and München.

Popper, Karl (1959): *The Logic of Scientific Discovery*. Translation of *Logik der Forschung*, Wien 1934 (translated by Karl Popper with the assistance of Julius Freed and Jan Freed). London.

Porter, Martin (1980): An Algorithm for Suffix Stripping. In: *Program* 14(3). pp. 130–137.

Porzig, Walter (1954): *Die Gliederung des indogermanischen Sprachgebiets* (= Indogermanische Bibliothek, series 6, vol. 4). Heidelberg.

Poultney, James Wilson (1963): Evidence for Indo-European Alternation of Initial /g^w/ and /w/. In: *Language* 39(3). pp. 398–408.

Prosdocimi, Aldo Luigi (1978): Diachrony and Reconstruction: 'genera proxima' and 'differentia specifica'. In: *Proceedings of the Twelfth International Congress of Linguists*, [Vienna 1977] (ed. by Wolfgang Dressler and Wolfgang Meid). Innsbruck. pp. 84–98.

Rankin, Robert Louis (2003): The Comparative Method. In: *The Handbook of Historical Linguistics* (ed. by Brian Daniel Jospeh and Richard David Janda). Malden. pp. 183–212.

Rasmussen, Jens Elmegård (1989): Die Tenues Aspiratae: Dreiteilung oder Vierteilung des indogermanischen Plosivsystems und die Konsequenzen dieser Frage für die Chronologie einer Glottalreihe. In: *The New Sound of Indo-European. Essays in Phonological Reconstruction. 7th International Conference on Historical Linguistics*, [Pavia 1985] (ed. by Theo Vennemann genannt Nierfeld). Berlin and New York City. pp. 153–176.

Rasmussen, Jens Elmegård (1994): On the Phonetics of the Indo-European Laryngeals. In: *In Honorem Holger Pedersen: Kolloquium der Indogermanischen Gesellschaft*, [Copenhagen 1993] (ed. by Jens Elmegård Rasmussen). Wiesbaden. pp. 433–447.

Rasmussen, Jens Elmegård (2009): Internal Reconstruction Applied to Indo-European: Where Do We Stand? In: *Internal Reconstruction in Indo-European. Methods, Results, and Problems. Section Papers from the XVI International Conference on Historical Linguistics*, [Copenhagen 2003] (ed. by Jens Elmegård Rasmussen and Thomas Olander). Copenhagen. pp. 255–268.

Rehn, Rudolf (1982): *Der logos der Seele. Wesen, Aufgabe und Bedeutung der Sprache in der platonischen Philosophie*. Hamburg.

Reichelt, Hans (1922): Die Labiovelare. In: *Indogermanische Forschungen* 40(1). pp. 40–81.

Ringe, Don (2003): Internal Reconstruction. In: *The Handbook of Historical Linguistics* (ed. by Brian Daniel Joseph and Richard David Janda). pp. 244–261.

Rix, Helmut (1992): *Historische Grammatik des Griechischen. Laut- und Formenlehre*. Revised 2nd edition. Darmstadt.

Roberts, Edward Arthur (2014): *A Comprehensive Etymological Dictionary of the Spanish Language with Families of Words Based on Indo-European Roots*, vol. 1: *A-G*. Bloomington/Indiana.

Rousseeuw, Peter (1987): Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. In: *Journal of Computational and Applied Mathematics* 20. pp. 53–65.

Sapir, Edward (1949): The Psychological Reality of Phonemes. In: *Selected Writings of Edward Sapir in Language, Culture, and Personality*. pp. 46–60. Translation of La Réalité psychologique des phonèmes. In: *Journal de Psychologie Normale et Pathologique* 30. pp. 247–265 (translated and edited by David Goodman Mandelbaum). Berkeley, Los Angeles, and London.

Saussure, Ferdinand de (1879): *Mémoire sur le système primitif des voyelles dans les langues indo-européennes*. Leipzig [reprint Hildesheim 1972].

Scheible, Silke / Schulte im Walde, Sabine & Springorum, Sylvia (2013): Uncovering Distributional Differences between Synonyms and Antonyms in a

Word Space Model. In: *Proceedings of the 6th International Joint Conference on Natural Language Processing,* [Nagoya 2013] (ed. by Ruslan Mitkov and Jong Park). Nagoya/Japan. pp. 489–497.

Schlegel, Friedrich (1808): *Ueber die Sprache und Weisheit der Indier. Ein Beitrag zur Begründung der Alterthumskunde. Nebst metrischen Uebersetzungen indischer Gedichte*. Heidelberg.

Schleicher, August (1868): Fabel in indogermanischer Ursprache. In: *Beiträge zur vergleichenden Sprachforschung auf dem Gebiete der arischen, celtischen und slawischen Sprachen* 5. pp. 206–208.

Schleicher, August (1869): *Indogermanische Chrestomathie. Schriftproben und lesestücke mit erklärenden glossaren zu August Schleichers compendium der vergleichenden grammatik der indogermanischen sprachen. Nebst zusätzen und berichtigungen zur zweiten auflage des compendiums.* (ed. by August Schleicher in collaboration with Hermann Ebel, August Leskien, and Johannes Schmidt). Weimar.

Schleicher, August (1871): *Compendium der vergleichenden Grammatik der indogermanischen Sprachen. Kurzer Abriss einer Laut- und Formenlehre der indogermanischen Ursprache, des Altindischen, Alteranischen, Altgriechischen, Altitalischen, Altkeltischen, Altslawischen, Litauischen und Altdeutschen*. Revised 3rd edition. Weimar.

Schmidt, Johannes (1872): Die Verwantschaftsverhältnisse der indogermanischen Sprachen. Weimar.

Schmidt, Karl Horst (1979): Reconstructing Active and Ergative Stages of Pre-Indo-European. In: *Ergativity. Towards a Theory of Grammatical Relations* (ed. by Frans Plank). London and New York City. pp. 333–345.

Schuler, Karin Kipper (2005): *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon* (Doctoral dissertation, University of Pennsylvania). Ann Arbor/Michigan.

Schultz, Julia (2012): *Twentieth Century Borrowings from French to English: Their Reception and Development*. Newcastle upon Tyne.

Schwartz, Roy / Reichart, Roi & Rappoport, Ari (2015): Symmetric Pattern Based Word Embeddings for Improved Word Similarity Prediction. *In Proceedings of the 19th Conference on Computational Language Learning*, [Beijing 2015] (ed. by Afra Alishahi and Alessandro Moschitti). Stroudsburg/Pennsylvania. pp. 258–267.

Schwyzer, Eduard & Debrunner, Albert (1939): *Griechische Grammatik*, vol 1: *Allgemeiner Teil, Lautlehre, Wortbildung, Flexion* (= Handbuch der Altertumswissenschaft 2). München.

Sedley, David (2003a): *Plato's Cratylus.* Cambridge/UK and New York City.

Sedley, David (2003b): The Nomothetês in Plato's Cratylus. In: *The Studia Philonica Annual* 15. pp. 5–16.

Seebold, Elmar (1980): Etymologie und Lautgesetz. In: Lautgeschichte und Etymologie. Akten der VI. Fachtagung der Indogermanischen Gesellschaft, [Vienna 1978] (ed. by Manfred Mayrhofer, Martin Peters, and Oskar Ernst Pfeiffer). Wiesbaden. pp. 431–484.

Selig, Maria (1992): *Die Entwicklung der Nominaldeterminanten im Spätlatein. Romanischer Sprachwandel und lateinische Schriftlichkeit* (= Script Oralia 26). Tübingen.

Sherzer, Joel (1972): Vowel Nasalization in Eastern Algonquian: An Areal-Typological Perspective on Linguistic Universals. In: *International Journal of American Linguistics* 38. pp. 267–268.

Shevelov, George Yurii (1964): *A Prehistory of Slavic. The Historical Phonology of Common Slavic*. Heidelberg [reprint New York City 1965].

Shimomiya, Tadao (1974): Entwicklung der zusammengesetzten Tempora als Phänomen des europäischen Sprachbundes. In: *Zeitschrift für vergleichende Sprachforschung* 88(2). pp. 218–229.

Sicoli, Mark & Holton, Gary (2014): Linguistic Phylogenies Support Back-Migration from Beringia to Asia. In: *PLoS One* 9(3). pp. 1–8.

Siebs, Theodor (1904): Anlautstudien. In: *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der indogermanischen Sprachen* 37(3). pp. 277–324.

Sigurd, Bengt (1966): Generative Grammar and Historical Linguistics. In: *Actra Linguistica Hafniensia* 10(1). pp. 35–48.

Sihler, Andreas Littleton (2000): *Language History* (= Amsterdam Studies in the Theory and History of Linguistic Science, series 4: Current Issues in Linguistic Theory, vol. 191). Amsterdam and Philadelphia.

Sims-Williams, Patrick (1998): The Celtic Languages. In: *The Indo-European Languages* (ed. by Anna Giacalone Ramta and Paoblo Ramat). pp. 345–379.

Sonderegger, Stefan (1979): *Grundzüge deutscher Sprachgeschichte. Diachronie des Sprachsystems*, vol. 1: *Einführung – Genealogie – Konstanten*. Berlin and New York City.

Spärck Jones, Karen (1972): A Statistical Interpretation of Term Specificity and Its Application in Retrieval. In: *Journal of Documentation* 28(5). pp. 11–21

Stanislav, Ján (1978): *Starosloviensky jazyk*, vol. 1: Veľká Morava a Panónia. Kultúrny jazyk a písomníctvo. *Konštantín Filozof, Metod a Kliment sloviensky. Fonetika*. Bratislava.

Starostin, George (2016): From Wordlists to Proto-Wordlists: Reconstruction as 'Optimal Selection'. In: *Faits de langues* 47(1). pp. 177–200.

Steiner, Lydia / Cysouw, Michael & Stadler, Peter (2011): A Pipeline for Computational Historical Linguistics. In: Language Dynamics and Change 1(1). pp. 89–127.

Stolz, Friedrich (1894): Zur Chronologie der lateinischen Lautgesetze. In: Indogermanische Forschungen 4. pp. 233–240.

Streitberg, Wilhelm (1896): *Urgermanische Grammatik. Einführung in das vergleichende Studium der altgermanischen Dialekte* (= Germanische Bibliothek, series 1, vol. 1). Heidelberg.

Sturtevant, Edgar Howard (1917): *Linguistic Change: An Introduction to the Historical Study of Language*. Chicago.

Sturtevant, Edgar Howard (1940): The Greek k-Perfect and Indo-European *k(o). In: *Language* 16(4). pp. 273–284.

Syrjänen, Kaj / Honkola, Terhi / Korhonen, Kalle / Lehtinen, Jyri / Vesakoski, Outi & Wahlberg, Niklas (2013): Shedding More Light on Language Classification Using Basic Vocabularies and Phylogenetic Methods: A Case Study of Uralic. In: *Diachronica* 30(3). pp. 323–352.

Szulc, Aleksander (1987): *Historische Phonologie des Deutschen.* (= Sprachstrukturen, series A: Historische Sprachstrukturen 6). Tübingen.

Tibshirani, Robert / Walther, Guenther & Hastie, Trevor (2001): Estimating the Number of Clusters in a Data Set via the Gap Statistic. In: *Journal of the*

Royal Statistical Society: Series B (Statistical Methodology) 63(2). pp. 411–423.

Tichy, Eva (2009): *Indogermanistisches Grundwissen für Studierende sprachwissenschaftlicher Disziplinen*. 3rd edition. Bremen.

Turney, Peter David & Pantel, Patrick (2010): From Frequency to Meaning: Vector Space Models of Semantics. In: *Journal of Artificial Intelligence Research* 37. pp. 141–188.

Unger, James Marshall (2000): Reconciling Comparative and Internal Reconstruction: The Case of Old Japanese / ti, ri, ni /. In: *Language* 76/3. pp. 655–681.

van Lessen Kloeke, Wus (1982): *Deutsche Phonologie und Morphologie. Merkmale und Markiertheit* (= Linguistische Arbeiten 117). Tübingen.

Vitali, David (2007): *Mit dem Latein am Ende? Volkssprachlicher Einfluss in lateinischen Chartularen aus der Westschweiz* (= Lateinische Sprache und Literatur des Mittelalters 41). Bern.

Wang, William Shi-Yuan (1969): Competing Sound Changes as a Cause of Residue. In: *Language* 45(1). pp. 9–25.

Waterman, Michael Spencer / Smith, Temple Ferris & Beyer, William (1976): Some Biological Sequences Metrics. In: *Advances in Mathematics* 20(3). pp. 367–387.

Watkins, Calvert (1978): Remarques sur la méthode de Ferdinand de Saussure comparatiste. In: *Cahiers Ferdinand de Saussure* 32. pp. 59–69.

Weiss, Sholom / Indurkhya, Nitin / Zhang, Tong / Damerau, Frederick Jacob (2010): *Text Mining. Predictive Methods for Analyzing Unstructured Information*. New York City.

Wells, Rulon Seymour (1949): Automatic Alternation. In: *Language* 25(2). pp. 99–116.

Werner, Otmar (1972): *Phonemik des Deutschen* (= Sammlung Metzler, series C, vol. 108). Stuttgart.

Wheeler, Ward & Whiteley, Peter Michael (2015): Historical Linguistics as a Sequence Optimization Problem: The Evolution and Biogeography of Uto-Aztecan Languages. In: *Cladistics* 31(2). pp. 113–125.

Wichmann, Søren / Holman, Eric & Brown, Cecil (2020): *The ASJP Database (version 19)* [page https://asjp.clld.org/]. Jena.

Wischer, Ilse (2011): Grammaticalization and Word Formation. In: *The Oxford Handbook of Grammaticalization* (edited by Heiko Narrog and Bernd Heine). Oxford. pp. 356–364.

Witting, Claes (1965): Somatische Bedingung der Aussprache und phonetische Methode. In: *STUF – Language Typology and Universals* 18(1–6). pp. 345–364.

Wodtko, Dagmar / Irslinger, Britta & Schneider, Carolin (2008): *Nomina im Indogermanischen Lexikon*. Heidelberg.

Yih, Wen-tau / Zweig, Geoffrey & Platt, John Carlton (2012): Polarity Inducing Latent Semantic Analysis. In: *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, [Jeju Island, Korea] (ed. by Jun'ichi Tsujii, James Henderson, and Marius Paşca). Stroudsburg/Pennsylvania. pp 1212–1222.

Zaliznjak, Anna Andreevna (1964): Синхронное описание и внутренняя реконструкция [Sinxronnoe opisanie i vnutrennjaja rekonstrukcija]. In: Проблемы сравительной грамматики индоевропейских языков. Научная сессия. Тезисы докладов [Problemy sravitel'noj grammatiki indoevropejskix jazykov. Nautschnaja sessija. Tezisy dokladov]. Moscow. pp. 51–54.

Zeps, Valdis Juris (1969): A Synchronic and Diachronic Order of Rules: Mutations of Velars in Old Church Slavonic. In: *Approaches in Linguistic Methodology* (ed. by Irmengard Rauch and Charles Thomas Scott). 2nd print. Madison/Wisconsin and Milwaukee. pp. 145–151.

Appendix A: Gold Standards

App. A.1 Gold Standard for German Sound Changes

The gold standard for the consonantal sound changes of German is based on Kümmel (2007:391–404). A crucial question is how to deal with regional or irregular sound changes. Rules classified as High German (ahd = Old High German, mhd = Middle High German, and nhd = New High German) or as common High German (g-ahd = common Old High German, g-mhd = common Middle High German, g-nhd = common New High German) are included. The time stages include sound changes that took place from the last language stage to the denoted language stage. The Pre-High German sound changes begin with Common Germanic (g-germ.). In exceptional cases, dialectal rules are also chosen if they had an influence on written High German. Dialects which are excluded from a sound change are denoted in curly brackets. For the meaning of the abbreviations, see Kümmel (2007).

The vocalic sound changes are taken from Paul (1969) and converted to formal notation. Sound changes that were regional and had no influence on written New High German are ignored, as are irregular changes (e.g., analogy or changes occurring only in particles, compounds, or certain morphemes). The development of unstressed vowels in Old High German is inconsistent since the rules can vary greatly depending on region and scribe (cf. Braune and Reiffenstein 2004:64–78). The chronological classification is disputable since sound changes often become visible much later in the written language. Irregular conditions are in brackets and are listed if they frequently occur.

(Pre-)Old High German Sound Changes

Single Sounds

x > h / _{-$,:}	g-germ.	x^{w} > h^{w} / _	g-germ.
ŋ > 0 / V>V:_h,x	g-germ.	g^{w} > gw / w	g-germ.
*z > *ɹ̯ / _	g-westgerm.	*ɹ̯ > 0 / V[-stress]_#	g-westgerm.
*ɹ̯ > r / _	g-westgerm.	ð > d / _	g-westgerm.
*θ > h / _lV	g-westgerm.	p > pp / _j	g-westgerm.
t > tt / _j	g-westgerm.	k > kk / _j	g-westgerm.
b > bb / _j	g-westgerm.	d > dd / _j	g-westgerm.
g > gg / _j	g-westgerm.	*φ > φφ / _j	g-westgerm.
*θ > θθ/ _j	g-westgerm.	x > xx / _j	g-westgerm.
s > ss / _j	g-westgerm.	*β > *ββ / _j	g-westgerm.
*ð > *ðð / _j	g-westgerm.	*γ > *γγ / _j	g-westgerm.
l > ll / _j	g-westgerm.	w > ww / _j	g-westgerm.
p > pp / _l,rV	g-westgerm.	t > tt / _l,rV	g-westgerm.
k > kk / _l,rV	g-westgerm.	x > xx / _l,rV	g-westgerm.
mm > m / V:, C_	g-westgerm. {-s-od.}	nn > n / V:, C_	g-westgerm. {-s-od.}
ll > l / V:, C_	g-westgerm. {-s-od.}	rr > r / V:, C_	g-westgerm. {-s-od.}
pp > p / V:, C_	g-westgerm. {-s-od.}	tt > t / V:, C_	g-westgerm. {-s-od.}
kk > k / V:, C_	g-westgerm. {-s-od.}	ss > s / V:, C_	g-westgerm. {-s-od.}
bb > b / V:, C_	g-westgerm. {-s-od.}	dd > d / V:, C_	g-westgerm. {-s-od.}
gg > g / V:, C_	g-westgerm. {-s-od.}	*φ> f / _	anfrk, g-ahd.
m > n / _φ>f	anfrk, g-ahd.	h > 0 / #_R	anfrk, g-ahd.
w > 0 / C_o,u	anfrk, g-ahd.	j > 0 / C{-r}_V	anfrk, g-ahd.
*x > χ / _$	anfrk, g-ahd.	k > x / s̲_	anfrk, g-ahd.

f > v̊ / #_	anfrk, g-ahd.	θ > ð̥ / #_	anfrk., g-ahd. {-lang.}
w > 0 / #_l	anfrk., g-ahd. {-mfrk.}	f > v / V_(R)V	anfrk., ahd.mfrk,n-th.
θ > ð / V_(R)V	anfrk., ahd.mfrk, n-th.	s̠ > z̠ / V_(R)V, #_	anfrk., ahd.mfrk, n-th.
v̊ > v / _	anfrk., ahd.mfrk, n-th.	*β > b / _	g-(vor)ahd.{-mfrk.,hess.,n-th.}
*ð > d / _	g-(vor)ahd.{-mfrk.,hess.,n-th.}	*γ > g / _	g-(vor)ahd.{-mfrk.,hess.,n-th.}
*p > p^h / {-s}_	*g-vorahd.	*t > t^h / {-s}_	*g-vorahd.
*k > k^h / {-s}_	*g-vorahd.	*p^h > pf / V_V	*g-vorahd.
*t^h > t̠s̠ / V_V	*g-vorahd.	*k^h > kx / V_V	*g-vorahd.
*θ > ð̥ / V(R)_	g-ahd.	t̠^{h} > t̠s̠ / #_	g-ahd.
pf > ff / V_V	g-ahd.	t̠s̠ > s̠s̠ / V_V	g-ahd.
kx > xx / V_V	g-ahd.	ð̥ð̥ (> d̥d̥) > tt / _	g-ahd.
j > j̧ / V̆r_	g-ahd.	ð̥ > d̥ / _	g-ahd.
x > ç / s̠_	g-ahd.	m > n / _#	g-ahd.
t > d̥ / n_	g-ahd.	pː > p / _#	g-ahd.
fː > f / _#	g-ahd.	tː > t / _#	g-ahd.
sː > s / _#	g-ahd.	kː > k / _#	g-ahd.
xː > x / _#	g-ahd.	ÞÞ > Þ / Vː_	ahd.{-s-od.}
j > r / r_	g-ahd.{-bair.}	bb > pp / _	g-ahd. {-mfrk., hess.,n-th.}
d > d̥ / _	g-ahd.{-mfrk.}	w > 0 / #_r	g-ahd.{-mfrk.}
d̥d̥ > tt / _	g-ahd.{-mfrk.}	*pf > f / l,r_	ahd.s-mfrk.,rhfrk.
h > 0 / V_V	ahd.dial.mhd. {-al.sz.dial., s-,m-bair.dial}	d̥ > t / _	ahd.od.
*p^h > pf / #_; _ː; C_	ahd.od.	k^h > kx / #_; _ː; C_	ahd.od.{-ofrk.}
g̊g̊ > kk / _	ahd.od.		

*e > *i / _NC, C$_0$i,j,u	(Paul § 19.1) Germ. > Early Germ.	*eu > *iu / _	Germ. > Pre-OHG
*ˈ_ē$_1$ > *ˈ_ā / _	Germ. > Pre-OHG	*ˈ_i > *[ˈ_ε] / _C$\{-NC\}$a,ā,e,ē, o,ō	(Paul § 19.1) Germ. > Pre-OHG
*ˈ_u > *[ˈ_ε] / _C$\{-NC\}$a,ā,e,ē, o,ō	(Paul § 19.1) Germ. > Pre-OHG	*ˈ_iu > *[ˈ_io] / _C$\left\{\begin{matrix} -NC \\ -w \end{matrix}\right\}$a,ā,e,ē, o,ō	(Paul § 19.1) Germ. > Pre-OHG
*j > i / C_#	(Braune / Heidermanns §119) Germ. > OHG	*w > o / C_#	(Braune / Heidermanns §119) Germ. > OHG
*ˈ_ē$_2$ > ˈ_ia / _	(Paul §17.1) Germ. > OHG	*ˈ_ō > ˈ_ua / _	(Paul §17.2) Germ. > OHG
*ai > ē / _r,w,h,#	(Paul §16a) Pre-OHG > OHG (7th century)	*au > ō / _h,[+alveolar]	(Paul §16b) Pre-OHG > OHG (8th century)
*ˈ_a > [ˈ_e] / _C$\left\{\begin{matrix} -hs,ht \\ -rw,rh \\ -lh \end{matrix}\right\}$i/ī/j	(Paul § 18, § 29.2) Pre-OHG > OHG (before 8th century)		

Transformed Sound Changes

x > h / _{-$,:}; x^w > h^w / _	g-germ.	ŋ > 0 / V>V:_h,x	g-germ.
g^w > gw / w	g-germ.	*z > *ɹ̝ > r/ _*	g-westgerm.
*ɹ̝ > 0 / V[-stress]_#	g-westgerm.	ð > d / _	g-westgerm.
*θ > h / _lV	g-westgerm.	T > TT ; N > NN ; l > ll ; w > ww / _j	g-westgerm.
p > pp ; t > tt; k > kk ; x > xx / _l,rV	g-westgerm.	CC > C / V:, C_	g-westgerm. {-s-od.}

*φ> f / _	anfrk, g-ahd.	m > n / _φ>f	anfrk, g-ahd.
h > 0 / #_R	anfrk, g-ahd.	w > 0 / C_o,u	anfrk, g-ahd.
j > 0 / C{-r}_V	anfrk, g-ahd.	*x > χ / _$	anfrk, g-ahd.
k > x / s̪_	anfrk, g-ahd.	f > v̊ / #_	anfrk, g-ahd.
θ > ð̥ / #_	anfrk., g-ahd.{-lang.}	w > 0 / #_l	anfrk., g-ahd. {-mfrk.}
f > v; θ > ð, s̪ > z̪ / V_(R)V	anfrk., ahd.mfrk,n-th.	s̪ > z̪ / #_	anfrk., ahd.mfrk,n-th.
v̊ > v / _	anfrk., ahd.mfrk,n-th.	*β > b, *ð > d, *γ > g / _	g-(vor)ahd.{-mfrk.,hess.,n-th.}
*p > p^h; *t > t^h; *k > k^h / {-s}_	*g-vorahd.	*p^h > pf; *t^h > t̪s̪; *k^h > kx / V_V	*g-vorahd.
*θ > ð̥ / V(R)_	g-ahd.	t̪h > t̪s̪ / #_	g-ahd.
pf > ff; t̪s̪ > s̪s̪; kx > xx / V_V	g-ahd.	ð̥ð̥ > d̥d̥ > tt / _	g-ahd.
j > ʝ / V̌r_	g-ahd.	ð̥ > d̥ / _	g-ahd.
x > ç / s̪_	g-ahd.	m > n / _#	g-ahd.
t > d̥ / n_	g-ahd.	Cː > C / _#	g-ahd.
ÞÞ > Þ / Vː_	ahd.{-s-od.}	j > r / r_	g-ahd.{-bair.}
d > d̥ > t/ _	g-ahd.{-mfrk.}	w > 0 / #_r	g-ahd.{-mfrk.}
bb > pp, d̥d̥ > tt, g̊g̊ > kk / _	g-ahd. {-mfrk.,hess.,n-th.}	*pf > f / l,r_	ahd.s-mfrk.,rhfrk.
h > 0 / V_V	ahd.dial.mhd.{-al.sz.dial.,s-,m-bair.dial}	*p^h > pf, k^h > kx / #_; _ː; C_	ahd.od.
*e > *i / _NC, C_0i,j,u	Germ. > Early Germ.	*e > *i / _NC, C_0i,j,u	Germ. > Early Germ.
*eu > *iu / _	Germ. > Pre-OHG	*eu > *iu / _	Germ. > Pre-OHG
*ˈē$_1$ > *ˈā / _	Germ. > Pre-OHG	*ˈē$_1$ > *ˈā / _	Germ. > Pre-OHG

*ai > ē / _r,w,h,#	Pre-OHG > OHG	*ai > ē / _r,w,h,#	Pre-OHG > OHG
*au > ō / _h,[+alveolar]	Pre-OHG > OHG	*au > ō / _h,[+alveolar]	Pre-OHG > OHG
*ˈa > [ˈe] / $_C\begin{Bmatrix}-hs, ht \\ -rw, rh \\ -lh\end{Bmatrix}$i/ī/j	Pre-OHG > OHG	*ˈa > [ˈe] / $_C\begin{Bmatrix}-hs, ht \\ -rw, rh \\ -lh\end{Bmatrix}$i/ī/j	Pre-OHG > OHG

$$ˈV\begin{Bmatrix}-long \\ +high \\ -low\end{Bmatrix} > ˈV\begin{Bmatrix}-long \\ -high \\ -low\end{Bmatrix} / \_C\{-NC\}V\begin{Bmatrix}-stress \\ -high\end{Bmatrix}$$

Germ. > Pre-OHG

$$V\begin{Bmatrix}-syllabic \\ +high \\ \alpha\ back\end{Bmatrix} > V\begin{Bmatrix}+syllabic \\ +high \\ \alpha\ back\end{Bmatrix} / C\_\#$$

Germ. > OHG

$$ˈV\!:\begin{Bmatrix}-high \\ -low \\ \alpha\ back\end{Bmatrix} > ˈV\begin{Bmatrix}+high \\ -low \\ \alpha\ back\end{Bmatrix}V\begin{Bmatrix}-high \\ +low \\ -syllabic\end{Bmatrix} / \_$$

Germ. > OHG

(Pre-)Middle High German Sound Changes

Single Sounds

b > p / _#	mhd.	d > t / _#	mhd.
g > k / _#	mhd.	v > f / _#	mhd.
ç > ʃ / s̪>ʃ_	mhd.	b > 0 / i,e_i	mhd.
d > 0 / i,e_i	mhd.	g > 0 / i,e_i	mhd.
r > 0 / V:_#	mhd.	t > k / _l	mhd.
d > g /_l	mhd.	w > β <w> / _	mhd.
t > ts / #_w	mhd.wmd.,od.	w > 0 / k_V	mhd.al.dial.
ua > uo / _	(Braune / Heidermanns §40) OHG	ˈa > [ˈæ] / _...i,ī,j,iu	(Paul §18, §29.3) OHG
_ˈo > _ˈö [_ˈœˀ] / _...i,ī,j,iu	(Paul §18) OHG	_ˈu > _ˈü [_ˈʏˀ] / _C{(-lt,ld,NC)} ...i,ī,j,iu	(Paul §18) OHG

_ˈā > _ˈæ [_ˈɛː] / _...i,ī,j,iu	(Paul §18) OHG	_ˈō > _ˈœ [_ˈœːˀ] / _...i,ī,j,iu	(Paul §18) OHG
ˈū > _ˈiu [ˈyː] / _...i,ī,j,iu	(Paul §18, §36) OHG	_ˈou > _ˈöu [_ˈœʏˀ] / _...i,ī,j,iu	(Paul §18) OHG
ˈuo > _ˈüe / _...i,ī,j,iu	(Paul §18) OHG	_ˈe [_ˈɛ] > _ˈe [_ˈe] / _...i,ī,j,iu	(Paul §18) OHG
ɛ > e / _sch,st	(Paul §18 fn. 6) OHG>MHG	ia,io > ie / _	OHG>MHG
a{-stress} > ə / _	(Paul §27) OHG>MHG	ā{-stress} > ə / _	(Paul §27) OHG>MHG
ɛ{-stress} > ə / _	(Paul §27) OHG>MHG	ē{-stress} > ə / _	(Paul §27) OHG>MHG
i{-stress} > ə / _	(Paul §27) OHG>MHG	ī{-stress} > ə / _	(Paul §27) OHG>MHG
o{-stress} > ə / _	(Paul §27) OHG>MHG	ō{-stress} > ə / _	(Paul §27) OHG>MHG
u{-stress} > ə / _	(Paul §27) OHG>MHG	ū{-stress} > ə / _	(Paul §27) OHG>MHG
ə > 0 / V[+stress]r,l , / C{α …}_C{α …} / V[+stress]C_0_C_0V[-stress] / h_s,t# and m_nt#			(Paul §24) MHG

Transformed Sound Changes

D > T / _#	mhd.	ç > ʃ / s̠>ʃ_	mhd.
b > 0; d > 0; g > 0 / i,e_i	mhd.	r > 0 / Vː_#	mhd.
t > k; d > g /_l	mhd.	w > β / _	mhd.
t > ts / #_w	mhd.wmd.,od.	w > 0 / k_V	mhd.al.dial.

ua > uo /	OHG	V{$-$ *stress*} > ə / _	OHG>MHG OHG

$$V\begin{Bmatrix} +stress \\ +back \\ \alpha\ high \\ \alpha\ low \\ \alpha\ long \\ \alpha\ diphthong \end{Bmatrix} > V\begin{Bmatrix} +stress \\ -back \\ \alpha\ high \\ \alpha\ low \\ \alpha\ long \\ \alpha\ diphthong \end{Bmatrix} /\ \_\ldots i, \bar{\imath}, j, iu$$

ə > 0 / V[+stress]r,l MHG
/ C{α …}_C{α …}
/ V[+stress]$$C_0$_$C_0$V[-stress]
/ h_s,t# and m_nt#

(Pre-)New High German Sound Changes

Single Sounds

n > 0 / V[+stress]_#	nhd.	h > 0 / l,r_	nhd.
g > ŋ / ŋ_	nhd.	m > n / V[-stress]_#	nhd.
w [β] > b / l,r_	g-nhd.{-dial.+}	ɉ > g / l,r_	g-nhd.{-dial.+}
j > 0 / V_V	g-nhd.{-dial.+}	w > 0 / V[+back]_V	g-nhd.{-dial.+}
s̺ > ʃ / r_ and #_C	g-nhd.{-dial.+}	z̺̊ > ʒ̊ / r_ and #_C	g-nhd.{-dial.+}
s̺ > s / X_V and #_V	g-nhd.{-dial.+}	z̺̊ > z / X_V and #_V	g-nhd.{-dial.+}
s̺ > s /_	g-nhd.{-dial.+}	ʁ > ɐ / _$	nhd.dial.
x > k / _s	nhd.md.,dial.,bair.s-al.sz.dial,w.,bas.	x > ç / {-V[-front]}_	nhd.md.,n-od.
b > m / m_V	nhd.od.{-dial.bair.s-al.}	kx > kh / _	
s̺ > ʃ / X_p	nhd.pf.,sfrk.ofrk,od.{-s-bair.z.7}	s̺ > ʃ / X_t	nhd.pf,sfrk,w-ostfrk.,al.,w-s-bair.,w-mbair.
r (> r̥) > ɐ / ə>0_	nhd.	m > m̥ / ə>0_	nhd.
n > n̥ / ə>0_	nhd.	ŋ > ŋ̩̊ / ə>0_	nhd.
l > l̥ / ə>0_	nhd.		

ə > 0 / #g_V,r,l,n,w	(Paul §24.7) MHG	ei > ai / _	(Paul §44) MHG > NHG

			(since 12^{th} century)
ou > au / _	(Paul §45) MHG>NHG (since 12^{th} century)	e > ö /_l,ʃ,C[+labial], C[+affricate]	(Paul §22a) MHG>NHG
aː > oː / (C[+nasal]), (C[+alveolar]), (C[+labial],h)	(Paul §22b) MHG>NHG	öu > äu/eu / _	(Paul §46) MHG>Early NHG
ie [ɪə̯ˀ] > je / #_	(Paul §30) MHG>NHG	æ > ɛ / _	MHG>NHG
æː > ɛː / _	MHG>NHG	a{+stress} > aː / _${-t,(-m,-r)}, _r#, _rC[+alveolar]	(Paul §23.1a.f,b,d) MHG>NHG (since 12^{th} century))
ɛ{+stress} > ɛː / _${-t,(-m,-r)},_r# ,_rC[+alveolar]	(Paul §23.1a.f,b,d) MHG>NHG	ɪ{+stress} > ɪ : / _${-t,(-m,-r)}, _r#	(Paul §23.1a.f,b) MHG>NHG
ɔ{+stress} > ɔː / _${-t,(-m,-r)}, _r#	(Paul §23.1a.f,b) MHG>NHG	ʊ{+stress} > ʊː / _${-t,(-m,-r)}, _r#	(Paul §23.1a.f,b) MHG>NHG
ü{+stress} > üː / _${-t,(-m,-r)}, _r#	(Paul §23.1a.f,b) MHG>NHG	ö{+stress} > öː / _${-t,(-m,-r)}, _r#	(Paul §23.1a.f,b) MHG>NHG
aː > a / (_ht, rC,) _CC, _t,er,x	(Paul §23.2 and §36) MHG>NHG	eː > e / (_ht, rC, CC)	(Paul §23.2) MHG>NHG
ɛː > ɛ / (_ht, rC, CC)	(Paul §23.2) MHG>NHG	iː > i / (_ht, rC, CC)	(Paul §23.2) MHG>NHG
oː > o / (_ht, rC, CC)	(Paul §23.2) MHG>NHG	uː > u / (_ht, rC, CC)	(Paul §23.2) MHG>NHG

ü: > ü / (_ht, rC, CC)	(Paul §23.2) MHG>NHG	ö: > ö / (_ht, rC, CC)	(Paul §23.2) MHG>NHG
i:{-stress}> ɪ / _	(Paul §26) MHG>NHG (since 13th century)	0 > ə / ī,ū,ʏ:_r(ə)#	(Paul §25) MHG>NHG
i: > ai / _	(Paul §20) MHG>NHG (12th–16th century)	u: > au / _	(Paul §20) MHG>NHG (12th–16th century)
ü: > oü / _	(Paul §20) MHG>NHG (12th–16th century)	ie > i: / _	(Paul §21) MHG>NHG (11th–12th century)
uo > u: / _	(Paul §21) MHG>NHG (11th–12th century)	üe > ʏ: / _	(Paul §21) MHG>NHG (11th–12th century)
u > o / _nn,mm,(n)	(Paul §32,35) MHG>NHG	ü > ö / _nn,mm,(n)	(Paul §32,35) MHG>NHG
ə > 0 / _#	(Paul §24.10) Early NHG>NHG	äu/eu > [ɔø] / _	(Paul §46) Early NHG>NHG
e > ɛ / _	(Paul §23.3) NHG	i > ɪ / _	(Paul §23.3) NHG
o > ɔ / _	(Paul §23.3) NHG	u > ʊ / _	(Paul §23.3) NHG
ø > œ / _	(Paul §23.3) NHG	y > ʏ / _	(Paul §23.3) NHG
ɛ: > e: / _	(Paul §23.3) NHG	ɪ: > i: / _	(Paul §23.3) NHG
ɔ: > o: / _	(Paul §23.3) NHG	ʊ: > u: / _	(Paul §23.3) NHG

œː > øː / _	(Paul §23.3) NHG	ʏː > yː / _	(Paul §23.3) NHG
ə > 0 / _ɐ	NHG		

Transformed Sound Changes

n > 0 / V[+stress]_#	nhd.	h > 0 / l,r_	nhd.
g > ŋ / ŋ_	nhd.	m > n / V[-stress]_#	nhd.
w [β] > b; ʝ > g / l,r_	g-nhd.{-dial.+}	j > 0 / V_V	g-nhd.{-dial.+}
Cː > C / _	g-nhd.{-dial.+}	w > 0 / V[+back]_V	g-nhd.{-dial.+}
s̺ > s ; z̺̊ > z / X_V und #_V	g-nhd.{-dial.+}	s̺ > ʃ ; z̺̊ > ʒ̊ / r_, #_C	g-nhd.{-dial.+}
ʁ > ɐ / _$	nhd.dial.	s̺ > s /_	g-nhd.{-dial.+}
x > ç; ɣ > ʝ / {-V[-front]}_	nhd.md.,n-od.	x > k / _s	nhd.md.,dial.,bair.s-al.sz.dial,w.,bas.
kx > kʰ / _	nhd.od.{-dial.bair.s-al.}	b > m / m_V	nhd.od.{-dial.bair.s-al.}
R > R̥ / ə>0_	nhd.	s̺ > ʃ / X_p, X_t	nhd.pf.,sfrk.,w-ofrk,,od.{-s-bair.z.7}

ə > 0 / #g_V,r,l,n,w	MHG	e > ö / _l,ʃ, C[+labial], C[+affricate]	MHG>NHG
aː > oː / (C[+nasal]), (C[+alveolar]), (C[+labial]),(h)	MHG>NHG	öu > äu/eu / _	MHG> Early NHG
ie [ɪə̯ˀ] > je / #_	MHG>NHG	iː{-stress}> ɪ / _	MHG>NHG
0 > ə / V{+high, +long}_r(ə)#	MHG>NHG	ə > 0 / _#	Early NHG>NHG
äu/eu > [ɔø] / _	Early NHG>NHG	ə > 0 / _ɐ	NHG

$$V\begin{Bmatrix} \alpha\, front \\ +\, syllabic \\ -\, high \\ -\, low \end{Bmatrix} V\begin{Bmatrix} \alpha\, front \\ -\, syllabic \\ +\, high \\ -\, low \end{Bmatrix} > V\begin{Bmatrix} +\, syllabic \\ -\, high \\ +\, low \end{Bmatrix} V\begin{Bmatrix} \alpha\, front \\ -\, syllabic \\ +\, high \\ -\, low \end{Bmatrix} /\_$$

MHG>NHG

$$V\begin{Bmatrix} +\, near\ open \\ \alpha\, long \end{Bmatrix} > V\begin{Bmatrix} -\, near\ open \\ \alpha\, long \end{Bmatrix}$$

MHG>NHG

$$V\begin{Bmatrix} +\, stress \\ -\, long \end{Bmatrix} > V\begin{Bmatrix} +\, stress \\ +\, long \end{Bmatrix} /\ \_\$\begin{Bmatrix} -\, t \\ -\, m \\ -\, er \end{Bmatrix}, \_r\#, \_rC[+alve.]$$

MHG>NHG

$$V\begin{Bmatrix} +stress \\ +\, long \end{Bmatrix} > V\begin{Bmatrix} +\, long \\ -\, long \end{Bmatrix} /\ \_ht,\ rC,\ \_CC$$

MHG>NHG

$$V\begin{Bmatrix} +\, high \\ +\, long \\ \alpha\, rounded \\ \alpha\, front \end{Bmatrix} > V\begin{Bmatrix} +\, syllabic \\ -\, high \\ +\, low \\ \alpha\, rounded \\ \alpha\, front \end{Bmatrix} V\begin{Bmatrix} -\, syllabic \\ +\, high \\ \alpha\, rounded \\ \alpha\, front \end{Bmatrix} /\_$$

MHG>NHG

$$V\begin{Bmatrix} +syllabic \\ +high \\ \alpha\, front \\ \alpha\, rounded \end{Bmatrix} V\begin{Bmatrix} -\, syllabic \\ -\, high \\ -\, low \\ \alpha\, rounded \\ \alpha\, front \end{Bmatrix} > V\begin{Bmatrix} +\, high \\ +\, long \\ \alpha\, rounded \\ \alpha\, front \end{Bmatrix} /\_$$

MHG>NHG

$$V\begin{Bmatrix} +syllabic \\ +high \\ \alpha\, front \\ \alpha\, rounded \end{Bmatrix} V\begin{Bmatrix} -\, syllabic \\ -\, high \\ -\, low \\ \alpha\, rounded \\ \alpha\, front \end{Bmatrix} > V\begin{Bmatrix} +\, high \\ +\, long \\ \alpha\, rounded \\ \alpha\, front \end{Bmatrix} /\_$$

MHG>NHG

$$V\begin{Bmatrix} +\, high \\ +\, rounded \end{Bmatrix} > V\begin{Bmatrix} -\, high \\ -low \\ +\, rounded \end{Bmatrix} /\ \_C\begin{Bmatrix} +\, nasal \\ +\, geminated \end{Bmatrix}$$

MHG>NHG

$$V\begin{Bmatrix} -\, long \\ -\, open \end{Bmatrix} > V\begin{Bmatrix} -\, long \\ +\, open \end{Bmatrix} /\_$$

NHG

$$V\begin{Bmatrix} +\, long \\ +\, open \end{Bmatrix} > V\begin{Bmatrix} +\, long \\ -\, open \end{Bmatrix} /\_$$

NHG

App. A.2 Gold Standard for German Phonological Rules

The gold standard for phonological rules in German is taken from O'Brien/Fagan (2006:114–124) and include only automatic rules (i.e., morphological alternations are excluded). The sound alternations listed under "Single Sounds" are all possible realizations of the rules listed under "Transformed Phonological Rules."

Single Sounds

aː > a / _[-stress]	eː > e / _[-stress]
ɛː > ɛ / _[-stress]	iː > i / _[-stress]
oː > o / _[-stress]	uː > u / _[-stress]
yː > y / _[-stress]	øː > ø / _[-stress]
ʒ > ʃ / _[-sonorant]$_0$.	b > p / _[-sonorant]$_0$.
d > t / _[-sonorant]$_0$.	g > k / _[-sonorant]$_0$.
z > s / _[-sonorant]$_0$.	v > f / _[-sonorant]$_0$.
ç > x / V[-front]_ *or* ç > χ / V[-front]_	g > ç / ɪ[-stress]_C_0.
0 > ʔ / $_V	ə > 0 / _C[+sonorant]C_0.
l > l̥ / .C_0_C_0.	ŋ > ŋ̥ / .C_0_C_0.
m > m̥ / .C_0_C_0.	n > n̥ / .C_0_C_0.
n > ŋ / [-continuant, -sonorant, +velar]_	n > m / [-continuant, -sonorant, +labial]_
r > ɐ / .C_0_C_0. *or* ʀ > ɐ / .C_0_C_0. *or* ʁ > ɐ / .C_0_C_0.	r > ɐ̯ / _C_0 *or* ʀ > ɐ̯ / _C_0 *or* ʁ > ɐ̯ / _C_0

Transformed Phonological Rules

V > [-long] / _[-stress]	Vowel shortening
[-sonorant] > [-voice] / _[-sonorant]$_0$.	Final devoicing
/ç/ > [-front] / V[-front]_	Dorsal fricative assimilation
/g/ > [ç] / $\begin{matrix}/\text{ɪ}/\\ [-stress]\end{matrix}$_$C_0$	Spirantization
Ø > [ʔ] / (_V…)$_\omega$	Glottal stop insertion

/ə/ > Ø / $\_\begin{bmatrix}\mathrm{C}\\ +\mathrm{sonorant}\end{bmatrix}$ C_0.	Schwa deletion
/R/ > [ɐ̯] / _C_0.	R-Vocalization
[+sonorant] > [+syllabic] / .C_0_C_0.	Sonorant Syllabification
$\begin{bmatrix}+nasal\\ +syllabic\end{bmatrix} > [\alpha\ place]\ /\ \begin{bmatrix}-continuant\\ -sonorant\\ \alpha\ place\end{bmatrix}$	Syllabic nasal assimilation

App. A.3 Gold Standard for Proto-Indo-European Phonological Rules

The phonological rules of the Proto-Indo-European gold standard are taken from Fortson (2010:69–72) and converted to formal notation. Bartholomae's and Osthoff's law are probably not of Proto-Indo-European date.

C > C[α voiced] / _C[α voiced]	Voicing assimilation
C[+ plosive] > C$\begin{bmatrix}+voiced\\ +aspirated\\ +plosive\end{bmatrix}$ / C[+plosive], s>C$\begin{bmatrix}+voiced\\ -aspirated\end{bmatrix}$_	Bartholomae's law
Ø > s / C[+dental]_ C[+ dental]	Dental-plus-dental clusters
s > Ø / s_	Simplification of *ss
C[+semivowel or +laryngeal] > Ø / V>V[+long]_C[+nasal]#	Stang's Law
s > Ø / V>V[+long]C[+resonant]s#	Szemerényi's Law
$C\begin{bmatrix}+velar\\ +labialized\end{bmatrix} > C\begin{bmatrix}+velar\\ -labialized\end{bmatrix} / u,w$	βουκόλος rule
C[α voiced] > C[- voiced] or C[+ voiced] / _#	Final voicing or devoicing
n > Ø / o[+long]_	Nasal loss

V[+long] > V[-short] / _C[+resonant]C	Osthoff's law
$\begin{bmatrix} +resonant\ or\ laryngeal \\ -syllabic \end{bmatrix} > \begin{bmatrix} +resonant\ or\ laryngeal \\ +syllabic \end{bmatrix}$ / {-syllabic}_ and $\begin{bmatrix} +resonant\ or\ laryngeal \\ +syllabic \end{bmatrix} > \begin{bmatrix} +resonant\ or\ laryngeal \\ -syllabic \end{bmatrix}$ / {+syllabic}_	Syllabification
$Ø > V \begin{bmatrix} +high \\ \alpha\ front \end{bmatrix} / .C\_C \begin{bmatrix} +semivowel \\ \alpha\ front \end{bmatrix} V$ *(final syllable)*	Sievers's and Lindeman's law

Appendix B: Paradigmatic Approach

App B.1 Result Paradigmatic Approach (Single Sounds)

Result list of the morphophonemic method (paradigmatic approach, single sounds) containing the 100 sound pairs with the highest Jaccard indices and the assignment to a phonological rule (PR), a Pre-New High German (NHG), a Pre-Middle High German (MHG), or a Pre-Old High German (OHG) sound change. Legend: PR = phonological rules, SC = sound change, FV = free variant, Abl = *ablauting* alternation, TS = transcription error, WC = wrong correspondence, R = sonorant, Sup = suppletive forms, I = i, ī, j, iu; →*NUMBER* = doublet of pair *NUMBER*.

	Pair	Jaccard		PR	SC	Example
1	ʀ|ɐ̯	0.779818	PR/ SC	r / ʀ / ʁ > ɐ̯ / C_0	ʁ > ɐ / _$ (NHG)	
2	l̩|l	0.739905	PR /SC	l > l̥ / .C_0_C_0.	l > l̥ / ə>0_ (NHG)	*plausible : plausibel*
3	ʁ|ɐ̯	0.702304	FV			
4	g|k	0.692642	PR/ SC	g > k / _[-R]$_0$.	g > k / _# (MHG)	
5	p|b	0.679333	PR/ SC	b > p / _[-R]$_0$.	b > p / _# (MHG)	
6	ə|ɐ	0.653292	PR/ SC	ə > 0 / C[+R]C_0.	ə > 0 / _ɐ (NHG)	
7	ə|l̩	0.651980	PR/ SC			→ 2 and 6
8	ʁ|ʀ	0.644229	FV			
9	ʀ|ɐ	0.642613	PR	r / ʀ / ʁ > ɐ / .C_0_C_0.		

10	ˈʊ|ʊ	0.640653	TS			
11	ˈɛː|ˈeː	0.630939	Abl			
12	ˈɪ|ˈɛ	0.629005	SC/ Abl		*e > *i / _NC, C_0...I (OHG)	*sprich:sprech schwind : schwänd*
13	z|s	0.623308	PR/ SC	z > s / _[-R]$_0$.	s̠ > s ; z̥̠ > z / X_V and #_V (NHG)	
14	ˈøː|ˈoː	0.622605	SC		ˈō > ˈœ / _...I (MHG)	
15	ˈɔ|ˈœ	0.616971	SC		ˈo > ˈö / _...I (MHG)	
16	ˈɛ|ɛ	0.615394	TS			
17	ˈyː|ˈaː	0.611980	Abl			
18	ˈiː|ˈaɪ̯	0.609059	Abl			
19	ˈuː|uː	0.606809	TS			
20	ˈɛː|ˈɔ	0.606496	Abl			
21	ˈyː|yː	0.603542	TS			
22	ˈaɪ̯|aɪ̯	0.600471	TS			
23	ˈa|a	0.598544	TS			
24	ˈaʊ̯|aʊ̯	0.594455	TS			
25	ˈɪ|ˈaɪ̯	0.590739	Abl			
26	ˈɔ|ɔ	0.586811	TS			
27	ˈøː|øː	0.571521	TS			
28	ˈɪ|ˈɔ	0.569511	Abl			
29	ˈɪ|ɪ	0.569441	TS			
30	ˈaː|aː	0.568078	TS			
31	ˈɛ|ˈa	0.562631	SC/ Abl		ˈa > [ˈæ] / _...I (MHG)	

32	g|ç	0.561831	PR	g > ç / ɪ[-stress]_C_0.		
33	t|d	0.559366	PR/ SC	d > t / _[-R]$_0$.	d > t / _# (MHG)	
34	ˈeː|eː	0.555258	TS			
35	ˈɔɪ̯|ɔɪ̯	0.554524	TS			
36	ˈiː|iː	0.553329	TS			
37	ˈɛ|ˈɔ	0.548434	Abl			
38	ˈøː|ˈiː	0.540942	Abl			
39	χ|x	0.535486	FV			
40	χ|ç	0.532577	PR/ SC	ç > x / V[-front]_	x > ç; ɣ > ʝ / {-V[-front]}_ (NHG)	
41	ˈiː|ˈeː	0.531996	SC			*seh:sieh gib:geb* (→ 12 + length-ening)
42	ˈoː|ˈiː	0.530663	Abl			
43	ˈʏ|ʏ	0.530615	TS			
44	ˈɛː|ɛː	0.525493	TS			
45	ʁ|ɐ	0.524262	FV			
46	ˈoː|oː	0.522889	TS			
47	ˈɛː|ˈiː	0.519109	Abl			*sieh:sähe*
48	ˈʊ|ˈɪ	0.514042	Abl			
49	ˈiː|ˈa	0.509717	Abl			
50	ˈɛː|ˈɛ	0.509601	TS			
51	ˈɪ|ˈeː	0.508761	SC			*nimm : neh-men* (→ 12 + lengthening)
52	ˈʊ|ˈɛ	0.493040	Abl			
53	ç|x	0.486893	FV			→ 40

54	ˈœ|ˈiː	0.486713	Abl			
55	ˈeː|ŋ	0.474824	Sup			*gehen : gegangen*
56	ˈɪ|ˈa	0.463950	Abl			
57	ˈœ|œ	0.456634	TS			
58	ç|k	0.455394	TS			
59	ˈɔ|ˈiː	0.453712	Abl			
60	ˈʏ|ˈɔ	0.448594	Abl			
61	ˈøː|ˈeː	0.444453	Abl			*heb:höb*
62	ˈoː|ˈeː	0.435250	Abl			
63	ˈɛ|ɪ	0.421859	SC/ Abl			*säng:sing*
64	ˈʏ|ˈɛ	0.419572	Abl			
65	ˈɪ|ˈœ	0.417898	Abl			
66	ŋ|ç	0.413951	SC		ŋ > 0 / V>V:_h,x (OHG)	*brächt:bring*
67	ˈʏ|ˈɪ	0.403524	Abl			
68	v|f	0.402418	PR	v > f / _[-R]$_0$.	v > f / _# (MHG)	
69	ˈɔ|ɪ	0.399594	Abl			
70	ˈiː|ˈaʊ̯	0.389439	Abl			
71	ˈɛ|a	0.381015	SC			
72	ˈɔ|ˈeː	0.379822	Abl			
73	n̩|n	0.377279	SC	n > n̥ / .C$_0$_C$_0$.	n > n̥ / ə>0_ (NHG)	*goldner : golden*
74	ˈɛː|ˈaː	0.374804	SC		ˈā > ˈæ / _...I (MHG)	
75	ˈeː|ˈa	0.373721	Abl			
76	ˈɛː|ˈoː	0.369121	Abl			

77	ˈøː|ˈyː	0.366174	Abl			
78	ˈyː|ˈa	0.365272	Abl			
79	ˈiː|ɡ	0.364522	WC			*ziehe:zogen*
80	ˈɛː|ɪ	0.364126	Abl			
81	ˈøː|oː	0.356435	SC			
82	ˈiː|ˈaː	0.353569	Abl			
83	ˈeː|d	0.352026	WC			
84	ˈʏ|ˈʊ	0.349671	SC		ˈu > ˈü / _C{-lt,-NC}...I (MHG)	
85	χ|ŋ	0.346873	SC			= 66
86	ˈyː|ˈoː	0.340402	Abl			
87	ˈɪ|aɪ̯	0.337389	Abl			
88	ˈɛː|t	0.336781	WC			
89	ˈɛ|ˈœ	0.336227	Abl			*schmelz : schmölz*
90	ˈiː|aɪ̯	0.331572	Abl			
91	ˈʏ|ˈeː	0.318438	Abl			
92	ˈæ|æ	0.315465	TS			
93	ˈøː|iː	0.314429	Abl			
94	ˈɛː|uː	0.312590	Abl			
95	ˈʊ|ɪ	0.309257	Abl			
96	ˈoː|iː	0.309020	Abl			
97	ˈeː|iː	0.306764	SC			
98	ˈɛ|ˈaː	0.305206	SC			*hätte : habe* (= 31 + shortening)
99	ˈuː|ˈiː	0.304245	Abl			
100	ɪ|ɛ	0.302326	TS			

App B.2 Phonological Rules (Paradigmatic, Single)

correctly identified	12	gold standard	26
free variants and doublets	10		
wrongly identified	78		
TOTAL	100		

Evaluation for the phonological rules using different threshold values. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	Correct Sound Pair	Number of Identified Pairs	Match Precision	Match Recall	Match F-Score
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-
0.8	0	0	-	-	-
0.75	1	1	1	0.038462	0.074074
0.7	1	1	1	0.076923	0.142857
0.65	*3*	*3*	*1*	*0.192308*	*0.322581*
0.6	2	14	0.368421	0.269231	0.311111
0.55	2	14	0.272727	0.346154	0.305085
0.5	1	11	0.227273	0.384615	0.285714
0.45	0	7	0.196078	0.384615	0.259740
0.4	1	9	0.183333	0.423077	0.255814
0.35	1	15	0.16	0.461538	0.237624
0.3	0	15	0.133333	0.461538	0.206897
0.25	-	-	0.133333	0.461538	0.206897

0.2	-	-	0.133333	0.461538	0.206897
0.15	-	-	0.133333	0.461538	0.206897
0.1	-	-	0.133333	0.461538	0.206897
0.0	-	-	0.133333	0.461538	0.206897
TOTAL	12	90			

App B.3 Old High German Sound Changes (Paradigmatic, Single)

Since Old High German			
correctly identified	17	gold standard	216
free variants and doublets	12	(including 2 PR)	
wrongly identified	71		
TOTAL	100		
Only Old High German			
correctly identified	2	gold standard	104
free variants and doublets	12		
wrongly identified	86		
TOTAL	100		

Evaluation for the sound changes since (Pre-)Old High German times using different threshold values. All sound changes listed in App. A.1 are considered. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	Correct Sound Pair	Number of Identified Pairs	Match Precision	Match Recall	Match F-Score
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-

0.8	0	0	-	-	-
0.75	1	1	1	0.009217	0.009217
0.7	1	1	1	0.018349	0.018349
0.65	3	3	1	0.045249	0.045249
0.6	4	13	0.5	0.076923	0.076923
0.55	2	13	0.354839	0.089069	0.089069
0.5	1	11	0.285714	0.093023	0.093023
0.45	0	7	0.244898	0.090566	0.090566
0.4	2	9	0.241379	0.102190	0.102190
0.35	2	15	0.219178	0.110727	0.110727
0.3	*1*	*15*	*0.193182*	*0.111842*	*0.111842*
0.25	-	-	0.193182	0.111842	0.111842
0.2	-	-	0.193182	0.111842	0.111842
0.15	-	-	0.193182	0.111842	0.111842
0.1	-	-	0.193182	0.111842	0.111842
0.0	-	-	0.193182	0.111842	0.111842
TOTAL	17	88			

App B.4 Middle High German Sound Changes (Paradigmatic, Single)

Since Middle High German			
correctly identified	15	gold standard	112
free variants and doublets	12	(including 2 PR)	
wrongly identified	73		
TOTAL	100		
Only Middle High German			
correctly identified	9	gold standard	37

free variants and doublets	12
wrongly identified	79
TOTAL	100

Evaluation for the sound changes since (Pre-)Middle High German times using different threshold values. All Pre-Middle and Pre-New High German sound changes listed in App. A.1 are considered. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	**Correct Sound Pair**	**Number of Identified Pairs**	**Match Precision**	**Match Recall**	**Match F-Score**
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-
0.8	0	0	-	-	-
0.75	1	1	1	0.008929	0.017699
0.7	1	1	1	0.017857	0.035088
0.65	3	3	1	0.044643	0.085470
0.6	3	13	0.444444	0.071429	0.123077
0.55	2	13	0.322581	0.089286	0.139860
0.5	1	11	0.261905	0.098214	0.142857
0.45	0	7	0.224490	0.098214	0.136646
0.4	1	9	0.206897	0.107143	0.141176
0.35	*2*	*15*	*0.191781*	*0.125*	*0.151351*
0.3	1	15	0.170455	0.133929	0.15
0.25	-	-	0.170455	0.133929	0.15
0.2	-	-	0.170455	0.133929	0.15
0.15	-	-	0.170455	0.133929	0.15
0.1	-	-	0.170455	0.133929	0.15
0.0	-	-	0.170455	0.133929	0.15
TOTAL	17	88			

App B.5 New High German Sound Changes (Paradigmatic, Single)

Since New High German			
correctly identified	6	gold standard	74
free variants and doublets	12	(including 2 PR)	
wrongly identified	82		
TOTAL	100		

Evaluation for the sound changes since (Pre-)New High German times using different threshold values. All (Pre)-New High German sound changes listed in App. A.1 are considered. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	Correct Sound Pair	Number of Identified Pairs	Match Precision	Match Recall	Match F-Score
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-
0.8	0	0	-	-	-
0.75	1	1	1	0.004630	0.009217
0.7	1	1	1	0.009259	0.018349
0.65	1	3	0.6	0.023148	0.045249
0.6	1	13	0.222222	0.041667	0.076923
0.55	0	13	0.129032	0.050926	0.089069
0.5	1	11	0.119048	0.055556	0.093023
0.45	0	7	0.102041	0.055556	0.090566
0.4	0	9	0.086207	0.064815	0.102190
0.35	1	15	0.082192	0.074074	0.110727

0.3	*0*	*15*	*0.068182*	*0.078704*	*0.111842*
0.25	-	-	0.068182	0.078704	0.111842
0.2	-	-	0.068182	0.078704	0.111842
0.15	-	-	0.068182	0.078704	0.111842
0.1	-	-	0.068182	0.078704	0.111842
0.0	-	-	0.068182	0.078704	0.111842
TOTAL	17	88			

App B.6 Correct Complementary Sounds (Paradigmatic, Single)

Evaluation of the correctly identified complementary sounds. Sound pairs whose sounds belong to a German phonological rule or sound change are marked in italics.

<table>
<tr><th colspan="3">Phonological Rules</th><th colspan="3">Sound Changes</th></tr>
<tr><th></th><th>Pair</th><th>Rule</th><th></th><th>Pair</th><th>Rule</th></tr>
<tr><td>1</td><td>ʀ|ɐ̯</td><td>r / ʀ / ʁ > ɐ̯ / _C_0</td><td>2</td><td>l̩|l</td><td>l > l̥ / ə>0_</td></tr>
<tr><td>2</td><td>l̩|l</td><td>l > l̥ / .C_0_C_0.</td><td>4</td><td>g|k</td><td>g > k / _#</td></tr>
<tr><td>4</td><td>g|k</td><td>g > k / _[-sonorant]$_0$.</td><td>5</td><td>p|b</td><td>b > p / _#</td></tr>
<tr><td>5</td><td>p|b</td><td>b > p / _[-sonorant]$_0$.</td><td>6</td><td>ə|ɐ</td><td>ə > 0 / _ɐ</td></tr>
<tr><td>6</td><td>ə|ɐ</td><td>ə > 0 / _C[+sonorant]C_0.</td><td>12</td><td>ˈɪ|ˈɛ</td><td>*e > *i / _NC, C_0i,j,u</td></tr>
<tr><td>9</td><td>ʀ|ɐ</td><td>r / ʀ / ʁ > ɐ / .C_0_C_0.</td><td>13</td><td>z|s</td><td>s̠ > s ; z̠̊ > z / X_V and #_V</td></tr>
<tr><td>13</td><td>z|s</td><td>z > s / _[-sonorant]$_0$.</td><td>14</td><td>ˈøː|ˈoː</td><td>ˈō > ˈœ [ˈœː?] / _...i,ī,j,iu</td></tr>
<tr><td>32</td><td>g|ç</td><td>g > ç / ɪ[-stress]_C_0.</td><td>15</td><td>ˈɔ|ˈœ</td><td>ˈo > ˈö [ˈœ?] / _...i,ī,j,iu</td></tr>
<tr><td>33</td><td>t|d</td><td>d > t / _[-sonorant]$_0$.</td><td>31</td><td>ˈɛ|ˈa</td><td>ˈa > [ˈæ] / _...i,ī,j,iu</td></tr>
<tr><td>53</td><td>x|ç</td><td>ç > x / V[-front]_</td><td>33</td><td>t|d</td><td>d > t / _#</td></tr>
</table>

<table>
<tr><td>*68*</td><td>*v|f*</td><td>*v > f / _[-sonorant]$_0$.*</td><td>*45*</td><td>*ʁ|ɐ*</td><td>*ʁ > ɐ / _$*</td></tr>
<tr><td>*73*</td><td>*n̩|n*</td><td>*n > n̥ / .C_0_C_0.*</td><td>*53*</td><td>*ç|x*</td><td>*x > ç; γ > ʝ / {-V[-front]}_*</td></tr>
<tr><td colspan="3" rowspan="5"></td><td>66</td><td>ŋ|ç</td><td>ŋ > 0 / V>V:_h,x</td></tr>
<tr><td>*68*</td><td>*v|f*</td><td>*v > f / _#*</td></tr>
<tr><td>*73*</td><td>*n̩|n*</td><td>*n > n̥ / ə>0_*</td></tr>
<tr><td>*74*</td><td>*ˈɛː|ˈaː*</td><td>*ˈā > ˈæ [ˈɛː] / _...i,ī,j,iu*</td></tr>
<tr><td>*84*</td><td>*ˈʏ|ˈʊ*</td><td>*ˈu > ˈü [ˈʏ?] / _C{-lt,ld,-NC}...i,ī,j,iu*</td></tr>
<tr><td colspan="2">TOTAL</td><td>11 (0.916667)</td><td>TOTAL</td><td colspan="2">14 (0.823529)</td></tr>
</table>

App B.7 Direction of the Phonological Rules (Paradigmatic, Single)

Evaluation of the correctly identified direction of German phonological rules. Sound pairs with correct direction are marked in italics. Column "C" stands for the most relevant condition. For the different methods, see Sect. 4.1.1.4.

<table>
<tr><th>Pair</th><th>Phonological Rule</th><th>C</th><th>Phonetic Plausibility</th><th>Articulatory Closer to the Sound Environment</th><th>Phonetic Distribution</th></tr>
<tr><td>1 ʀ|ɐ̯</td><td>r,ʀ,ʁ > ɐ̯ / _C_0</td><td>_ə</td><td>*ʀ>ɐ*</td><td>*ʀ>ɐ̯*</td><td>ɐ̯>ʀ</td></tr>
<tr><td>2 l̩|l</td><td>l > l̥ / .C_0_C_0.</td><td>_t</td><td>-</td><td>l̩>l</td><td>*l>l̩*</td></tr>
<tr><td>4 g|k</td><td>g > k / _[-R]$_0$.</td><td>_ə</td><td>-</td><td>k>g</td><td>k>g</td></tr>
<tr><td>5 p|b</td><td>b > p / _[-R]$_0$.</td><td>_t</td><td>-</td><td>*b>p*</td><td>*b>p*</td></tr>
<tr><td>6 ə|ɐ</td><td>ə > 0 / _C[+R]C_0.</td><td>_ʀ</td><td>-</td><td>-</td><td>ɐ>ə</td></tr>
<tr><td>9 ʀ|ɐ</td><td>r,ʀ,ʁ > ɐ / .C_0_C_0.</td><td>_ə</td><td>*ʀ>ɐ*</td><td>*ʀ>ɐ*</td><td>ɐ>ʀ</td></tr>
</table>

13 z|s	z > s / _[-R]$_0$.	_ə	-	s>z	s>z
32 g|ç	g > ç / ɪ[-stress]_C_0.	_ə	-	ç>g	ç>g
33 t|d	d > t / _[-R]$_0$.	_#	-	*d>t*	*d>t*
53 ç|x	ç > x / V[-front]_	ˈɛ	-	x>ç	x>ç
68 v|f	v > f / _[-R]$_0$.	_ə	-	f>v	f>v
73 n̩|n	n > n̥ / .C_0_C_0.	n̩	-	*n>n̩*	*n>n̩*
TOTAL			2	5	4
			0.166667	0.416667	0.333333

App B.8 Direction of the Sound Changes (Paradigmatic, Single Sounds)

Evaluation of the correctly identified direction of German sound changes. Sound pairs with correct direction are marked in italics. Column "C" stands for the most relevant condition. For the different methods, see Sect. 4.1.1.4.

Pair	Phonological Rule	C	Phonetic Plausibility	Articulatory Closer to the Sound Environment	Phonetic Distribution
2 l̩|l	l > l̥ / ə>0_	_t	-	l̩>l	*l>l̩*
4 g|k	g > k / _#	_ə	-	k>g	k>g
5 p|b	b > p / _#	_t	-	*b>p*	*b>p*
6 ə|ɐ	ə > 0 / _ɐ	_ʀ	-	-	ɐ>ə
12 ˈɪ|ˈɛ	*e > *i / _NC,C_0i,j,u	z_t͡s, _t͡s	-	*ˈɛ> ˈɪ*	*ˈɛ> ˈɪ*
13 z|s	z > s / _#	_ə	-	s>z	s>z
14 ˈøː|ˈoː	ˈō > ˈœ [ˈœː?] / _...i,ī,j,iu	-	-	-	ˈoː>ˈøː

15 ˈɔ|ˈœ	ˈo > ˈö [ˈœ?] / _...i,ī,j,iu	n̩	-	-	ˈœ>ˈɔ
31 ˈɛ|ˈa	ˈa > [ˈæ] / _...i,ī,j,iu	_ç	-	*ˈa>ˈɛ*	*ˈa>ˈɛ*
33 t|d	d > t / _#	_#	-	*d>t*	*d>t*
45 ʁ|ɐ	ʁ > ɐ / _$	_ə	ʀ>ɐ	*ʀ>ɐ*	ɐ>ʀ
53 ç|x	x > ç; γ > ʝ / {-V[-front]}_	ˈɛ	-	*x>ç*	ç>x
66 ŋ|ç	ŋ > 0 / V>V:_h,x	ˈɪ, ˈɪ_	-	ŋ>ç	ç>ŋ
68 v|f	v > f / _#	_ə	-	f>v	f>v
73 n̩|n	n > n̥ / ə>0_	n̩	-	*n>n̩*	*n>n̩*
74 ˈɛː|ˈaː	ˈā > ˈæ [ˈɛː] / _...i,ī,j,iu	s_s	-	*ˈaː>ˈɛː*	*ˈaː>ˈɛː*
84 ˈʏ|ˈʊ	ˈu > ˈü [ˈʏ?] / _C{-lt,ld,-NC}...i,ī,j,iu	-	-	-	*ˈʊ>ˈʏ*
TOTAL			1	8	8
			0.058824	0.470588	0.470588

App B.9 Conditions of the Phonological Rules (Paradigmatic, Single Sounds)

Evaluation of the identified conditions of German phonological rules. If the direction was determined incorrectly, the condition was adjusted (column "Adj. Rule"). For more details, see Sect. 7.5.1.4. Legend: P = number of identified conditions (threshold > 0.9, see Sect. 8.1.3 and 8.2.4), TP = true positives.

Pair	Phonological Rule	Adj. Rule	P	TP	Preci-sion	Doublets
1 ʀ|ɐ̯	r,ʀ,ʁ > ɐ̯ / _C_0	_V	46	34	0.739130	0.970588
2 l̥|l	l > l̥ / .C_0_C_0.	-	34	34	1	0.970588
4 g|k	g > k / _[-R]$_0$.	_V, _C[+voiced]	22	21	0.954545	0.952381

5 p|b	b > p / _[-R]$_0$.	-	35	29	0.828571	0.965517
6 ə|ɐ	ə > 0 / _C[+R]C$_0$.	_/r,ʁ,ʀ/	15	15	1	0.866667
9 ʀ|ɐ	r,ʀ,ʁ > ɐ / .C$_0$_C$_0$.	_V	7	5	0.714286	0.8
13 z|s	z > s / _[-R]$_0$.	_V, _C[+voiced]	19	19	1	0.894737
32 g|ç	g > ç / ɪ[-stress]_C$_0$.	_V	2	2	1	0.5
33 t|d	d > t / _[-R]$_0$.	-	18	10	0.555556	0.8
53 ç|x	ç > x / V[-front]_	V[front]_, C_	11	10	0.909091	0.8
68 v|f	v > f / _[-R]$_0$.	_V, _C[+voiced]	6	6	1	0.666667
73 n̥|n	n > n̥ / .C$_0$_C$_0$.	-	1	0	0	-
TOTAL			216	185	0.856481	0,835195

App B.10 Conditions of the Sound Changes (Paradigmatic, Single Sounds)

Evaluation of the identified conditions of German sound changes. If the direction was determined incorrectly, the condition was adjusted (column "Adj. Rule"). For more details, see Sect. 7.5.1.4. Legend: P = number of identified conditions (threshold > 0.9, see Sect. 8.1.3 and 8.2.4), TP = true positives.

Pair	Sound Change	Adj. Rule	P	TP	Precision	Doublets
2 l̥|l	l > l̥ / ə>0_	-	34	34	1	0.970588
4 g|k	g > k / _#	-	22	21	0.954545	0.952381
5 p|b	b > p / _#	-	35	29	0.828571	0.965517
6 ə|ɐ	ə > 0 / _ɐ	-	15	15	1	0
12 ˈɪ|ˈɛ	*e > *i / _NC, C$_0$i,j,u	-	2	0	0	

13 z|s	z > s / _#	-	19	19	1	0.894737
14 ˈøː|ˈoː	ˈō > ˈœ [ˈœː?] / _...i,ī,j,iu	-	0	0	-	-
15 ˈɔ|ˈœ	ˈo > ˈö [ˈœ?] / _...i,ī,j,iu	-	3	0	0	0
31 ˈɛ|ˈa	ˈa > [ˈæ] / _...i,ī,j,iu	-	5	0	0	0
33 t|d	d > t / _#	-	18	10	0.555556	0.8
45 ʁ|ɐ	ʁ > ɐ / _$	_V	3	2	0.666667	0.5
53 ç|x	x > ç; ɣ > ʝ / {-V[-front]}_	-	11	10	0.909091	0.8
66 ŋ|ç	ŋ > 0 / V>Vː_h,x	-	10	0	0	0
68 v|f	v > f / _#	-	6	6	1	0.666667
73 n̩|n	n > n̥ / ə>0_	-	1	0	0	0
74 ˈɛː|ˈaː	ˈā > ˈæ [ˈɛː] / _...i,ī,j,iu	-	1	0	0	0
84 ˈʏ|ˈʊ	ˈu > ˈü [ˈʏ?] / _C{-lt,ld,-NC}...i,ī,j,iu	-	0	0	-	-
TOTAL			185	146	0.789189189	0.72776555
only diachronic sound changes			32	10	0.3125	

App B.11 Result Paradigmatic Approach (Transformed Rules)

Result list of the morphophonemic method (paradigmatic approach, transformed rules) containing the 100 sound pairs with the highest Jaccard indices and the assignment to a phonological rule (PR), a Pre-New High German (NHG), a Pre-Middle High German (MHG), or a Pre-Old High German (OHG) sound change. Legend: PR = phonological rules, SC = sound change, FV = free variant, Abl = *ablauting* alternation, TS = transcription error, WC = wrong correspondence, R = sonorant, Sup = suppletive forms, I = i, ī, j, iu; →*NUMBER* = doublet of pair *NUMBER*.

Pair	Jaccard		Sound Pairs	PR	SC
1 vibrant-central vibrant-near_open	0.755414	PR/ SC	ʀ-ɐ̯, ʀ-ɐ	r,ʁ,ʀ > ɐ / _C_0	ʁ > ɐ / _$ (NHG)
1 uvular-near_open	0.678119	→1	ʁ-ɐ̯, ʁ-ɐ		-
2 mid-near_open	0.653292	PR/ SC	ə-ɐ	/ə/ > Ø / _C[+R]C_0	ə > 0 / _ɐ or r̥ > ɐ (NHG)
2 short-lateral mid-lateral mid-approximant	0.651980	→3	ə-l̩		R > R̥ / ə>0_ (NHG)
3 voiceless-voiced	0.610616	PR	g-k, p-b, z-s, g-ç, t-d, ŋ-ç, v-f, χ-ŋ, ŋ-x, ʀ-s, p-m̩, χ-g, …	[-R] > [-voiced] / _[-R]	contains s̯ > s ; z̯̊ > z / X_V and #_V (NHG), D > T / _# (MHG)
-	-	FV	ʁ-ʀ		
4 stressed-unstres-sed	0.590648	TS	ˈʊ-ʊ, ˈɛ-ɛ, ˈuː-uː, ˈyː-yː, ˈaɪ̯-aɪ̯, ˈa-a, ˈɔ-ɔ, ˈøː-øː, …		
5 close-unrounded	0.599788	Abl	ˈyː-ˈaː, ˈiː-ˈaɪ̯, ˈyː-ˈa, ˈɛː-ˈuː,		

			iː-aɪ̯, ˈuː-ˈaː		
6 front-back	0.597483	SC	ˈøː-ˈoː, ˈɔ-ˈœ, …		*umlaut* (MHG)
7 mid_close-close	0.561221	Abl	ˈøː-ˈiː, ˈiː-ˈeː, ˈøː-ˈyː, iː-eː, ˈuː-ˈoː		
8 near_close-open_mid centralized-open_mid	0.558259	SC/ Abl	ˈɪ-ˈɛ, ˈɪ-ˈɛː, ˈʏ-ˈɛ, ˈɪ-ˈœ, ɪ-ɛ, ˈɔ-ˈaʊ̯, ˈɔ-aʊ̯, ɪ-œ, ɛː-aɪ̯, ɪ-ɛː		*e > *i / _NC, C_0i,j,u (OHG)
9 short-unrounded	0.566305	Abl	ˈɪ-ˈaɪ̯, ˈœ-ˈiː, ˈʏ-ˈeː, ɔɪ̯-ɐ		
10 uvular-palatal	0.532577	SC	χ-ç	/ç/ > [-front] / V[-front]_	x > ç / {-V[-front]}_ (NHG)
11 short-long	0.506110	Abl / SC/ TS	ˈiː-ˈa, ˈɛː-ˈɛ, ˈɪ-ˈeː, ˈeː-ˈa, ˈɛ-ˈaː, ˈɛ-ˈeː, ˈaː-ˈa, ˈʊ-ˈuː, …	V > [-long] / _[-stress]	V[+accented, +long] > V[+accented, -long] /_ht,rC (MHG)

12 -	0.525220	Abl	ˈɛː-ˈeː	-	
13 open_mid-open	0.515784	→9	ˈɛ-ˈa, ˈɛː-ˈaː, ɛ-a, ˈœ-ˈa		
14 palatal-velar	0.505380	TS	ç-x, ç-k		
15 -	0.469954	Abl	ˈɛː-ˈiː		
16 rounded-un-rounded	0.460854	Abl	ˈøː-ˈeː, ˈʏ-ˈɪ, ˈɛ-ˈœ, ˈʏ-ˈa, ˈɛː-ˈøː		
18 -	0.430417	Abl	ˈɪ-ˈa		
19 long-velar	0.423795	SC/ Supp	ˈeː-ŋ, ˈiː-g, ˈɛː-g, ŋ-eː, g-iː, ˈɔɪ̯-ŋ, g-eː		ŋ > 0 / V>V:_h,x (OHG)
20 plosive-long	0.389231	-	ˈeː-d, ˈɛː-t, ˈøː-t, ˈiː-t, ʔ-yː, ʔ-aʊ̯, b-aɪ̯, ʔ-aɪ̯, ʔ-uː, …		
-	0.388272	FV	χ-x		
21 close-open	0.353569	Abl	ˈiː-ˈaː		
22 affricate-frica-tive	0.347179	SC	t͡s-s, p͡f-f, x-t͡ʃ		

23 vowel-voiced	0.298414	-	ˈa-ŋ, ˈuː-n, ˈɛː-z, n-aɪ̯, ˈɛː-n, ə-m̩, ə-ŋ, ʀ-aɪ̯, ə-n̩, …		
24 affricate-plosive	0.277536	TS	p͡f-p, t͡s-t, t͡ʃ-t		
25 fricative-plosive	0.238928	TS	χ-k, ç-t, x-k, ç-p, t-s, ɡ-v, ʁ-ɡ		
26 central-mid_close mid-front mid-unrounded mid-mid_close	0.238340	Abl	ə-e		
27 -	0.220563	Abl	ˈoː-ˈaʊ̯		
28 centralized-close near_close-close	0.190575	SC	ɪ-i, ˈɔɪ̯-ˈiː		*z > *ɹ̠ / _ (OHG)
29 -	0.178791	-	ʀ-z		
30 -	0.150994	-	ʃ-s		
31 alveolar-labio-dental	0.160963	-	z-v		
32 -	0.146680	TS	t͡s-t		
33 -	0.088615	Abl	ɔ-o		
34 mid_close-open	0.083645	Abl	ˈeː-ˈaː		

App B.12 Phonological Rules (Paradigmatic, Transformed Rules)

correctly identified	5	gold standard	9
free variants and doublets	4		
wrongly identified	25		
TOTAL	34		

Evaluation for the phonological rules using different threshold values. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	Correct Sound Pair	Number of Identified Pairs	Match Precision	Match Recall	Match F-Score
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-
0.8	0	0	-	-	-
0.75	1	1	1	0.111111	0.2
0.7	0	0	1	0.111111	0.2
0.65	1	1	1	0.222222	0.363636
0.6	*1*	*1*	*1*	*0.333333*	*0.5*
0.55	0	6	0.333333	0.333333	0.333333
0.5	2	4	0.384615	0.555556	0.454545
0.45	0	3	0.3125	0.555556	0.4
0.4	0	2	0.277778	0.555556	0.370370
0.35	0	2	0.25	0.555556	0.344828
0.3	0	1	0.238095	0.555556	0.333333
0.25	0	2	0.217391	0.555556	0.3125
0.2	0	4	0.185185	0.555556	0.277778
0.15	0	3	0.166667	0.555556	0.256410
0.1	0	0	0.166667	0.555556	0.256410
0.0	0	0	0.166667	0.555556	0.256410
TOTAL	5	30			

App B.13 Old High German Sound Changes (Paradigmatic, Transformed)

Since Old High German			
correctly identified	9	gold standard	104
free variants and doublets	5		
wrongly identified	20		
TOTAL	34		
Only Old High German			
correctly identified	4	gold standard	54
free variants and doublets	5		
wrongly identified	25		
TOTAL	34		

Evaluation for the sound changes since (Pre-)Old High German times using different threshold values. All sound changes listed in App. A.1 are considered. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	Correct Sound Pair	Number of Identified Pairs	Match Precision	Match Recall	Match F-Score
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-
0.8	0	0	-	-	-
0.75	1	1	1	0.026316	0.051282
0.7	0	0	1	0.026316	0.051282
0.65	1	1	1	0.052632	0.1
0.6	0	1	0.666667	0.052632	0.097561
0.55	0	5	0.25	0.052632	0.086957

0.5	*2*	*4*	*0.333333*	*0.105263*	*0.16*
0.45	0	3	0.266667	0.105263	0.150943
0.4	0	2	0.235294	0.105263	0.145455
0.35	0	2	0.210526	0.105263	0.140351
0.3	0	1	0.2	0.105263	0.137931
0.25	0	2	0.181818	0.105263	0.133333
0.2	0	4	0.153846	0.105263	0.125
0.15	0	3	0.137931	0.105263	0.119403
0.1	0	0	0.137931	0.105263	0.119403
0.0	0	0	0.137931	0.105263	0.119403
TOTAL	4	29			

App B.14 Middle High German Sound Changes (Paradigmatic, Transformed Rules)

Since Middle High German			
correctly identified	6	gold standard	50
free variants and doublets	5		
wrongly identified	23		
TOTAL	34		
Only Middle High German			
correctly identified	3	gold standard	12
free variants and doublets	5		
wrongly identified	26		
TOTAL	34		

Evaluation for the sound changes since (Pre-)Middle High German times using different threshold values. All Pre-Middle and Pre-New High German sound changes listed in App. A.1 are considered. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	**Correct Sound Pair**	**Number of Identified Pairs**	**Match Precision**	**Match Recall**	**Match F-Score**
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-
0.8	0	0	-	-	-
0.75	1	1	1	0.020408	0.04
0.7	0	0	1	0.020408	0.04
0.65	1	1	1	0.040816	0.078431
0.6	1	1	1	0.061224	0.115385
0.55	1	6	0.444444	0.081633	0.137931
0.5	*2*	*3*	*0.5*	*0.122449*	*0.196721*
0.45	0	3	0.4	0.122449	0.1875
0.4	0	2	0.352941	0.122449	0.181818
0.35	0	2	0.315789	0.122449	0.176471
0.3	0	1	0.3	0.122449	0.173913
0.25	0	2	0.272727	0.122449	0.169014
0.2	0	4	0.230769	0.122449	0.16
0.15	0	3	0.206897	0.122449	0.153846
0.1	0	0	0.206897	0.122449	0.153846
0.0	0	0	0.206897	0.122449	0.153846
TOTAL	6	29			

App B.15 New High German Sound Changes (Paradigmatic, Transformed Rules)

Since New High German			
correctly identified	4	gold standard	38
free variants and doublets	5		
wrongly identified	25		
TOTAL	34		

Evaluation for the sound changes since (Pre-)New High German times using different threshold values. All (Pre)-New High German sound changes listed in App. A.1 are considered. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	Correct Sound Pair	Number of Identified Pairs	Match Precision	Match Recall	Match F-Score
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-
0.8	0	0	-	-	-
0.75	1	1	1	0.026316	0.051282
0.7	0	0	1	0.026316	0.051282
0.65	1	1	1	0.052632	0.1
0.6	0	1	0.666667	0.052632	0.097561
0.55	0	5	0.25	0.052632	0.086957
0.5	2	4	0.333333	0.105263	0.16
0.45	*0*	*3*	*0.266667*	*0.105263*	*0.150943*
0.4	0	2	0.235294	0.105263	0.145455
0.35	0	2	0.210526	0.105263	0.140351

0.3	0	1	0.2	0.105263	0.137931
0.25	0	2	0.181818	0.105263	0.133333
0.2	0	4	0.153846	0.105263	0.125
0.15	0	3	0.137931	0.105263	0.119403
0.1	0	0	0.137931	0.105263	0.119403
0.0	0	0	0.137931	0.105263	0.119403
TOTAL	4	29			

App B.16 Correct Complementary Sounds (Paradigmatic, Transformed Rules)

Evaluation of the correctly identified complementary sounds. Feature pairs whose sounds belong to a German phonological rule or sound change are marked in italics. The first table shows the result for phonological rules, the second table for sound changes. The column “P” stands for the positives, “T” for true positives.

Feature Pair	**Phonological Rule**	**P**	**T**	**Preci-sion**	**Recall**	**F-Score**
1 vibrant-central, vibrant-near_open	r,ʁ,ʀ > ɐ / _C_0	2	1	0.5	1	0.66667
3 mid-near_open	/ə/ > Ø / _C[+R]C_0	1	0	0	0	-
5 voiceless-voiced	[-R]>[-voice] / _[-R]	24	6	0.25	0.85714	0.38710
13 χ-ç	/ç/ > [-front] / V[-front]_	1	1	1	1	1
14 short-long	V > [-long] / _[-stress]	34	5	0.14706	0.55556	0.23256
TOTAL		62	13	0.20968	0.56522	0.30588

Feature Pair	Sound Change	P	T	Preci-sion	Recall	F-Score
1 vibrant-central vibrant-near_open	ʁ > ɐ / _$	2	1	0.5	1	0.66667
3 mid-near_open	R>R̥ / ə>0_	1	0	0	0	-
5 voiceless-voiced	D > T / _#	24	6	0.25	0.857143	0.38710
9 front-back	*umlaut*	34	5	0.147059	0.555556	0.23256
11 near_close-open_mid centralized-open_mid	*e > *i / _NC, C_0i,j,u	10	1	0.1	1	0.18182
13 χ-ç	x > ç / {-V[-front]}_	0	0	0	0	-
14 short-long	V[+stress, -long] > V[+stress, +long] /_...	34	5	0.147059	0.555556	0.23256
21 long-velar	ŋ > 0 / V>V:_h,x	7	0	0	0	
34 ʀ-z	*z > *ɹ̝ / _	1	1	1	1	1
TOTAL		113	19	0.166667	0.542857	0.25503

App B.17 Direction of the Phonological Rules (Paradigmatic, Transformed)

Evaluation of the correctly identified direction of German phonological rules. Feature pairs with correct direction are marked in italics. Column “C” stands for the most relevant condition; “AA” for *alveolar_affricate*. For the different methods, see Sect. 4.1.1.4.

Feature Pair	Phonological Rule	C	Phonetic Plausibility	Articulatory Closer to the Sound Environment	Phonetic Distribution
1 vibrant-central, vibrant-near_open	r,ʁ,ʀ > ɐ / _C_0	ˈiː_ə	vibrant > central / near_open	vibrant > central / near_open	central / near_open > vibrant
2 mid-near_open	/ə/ > Ø / _C[+R]C_0	_ʀ	-	-	near_open > mid
5 voiceless-voiced	[-R]>[-voice] / _[-R]	ˈeː_t	-	voiced > voiceless	voiced > voiceless
13 χ-ç	ç > [-front] / V[-front]_	_#	-	-	fricative > palatal
14 short-long	V > [-long] / _[-stress]	AA	-	-	long > short
TOTAL			1	2	2
			0.2	0.4	0.4

App B.18 Direction of the Sound Changes (Paradigmatic, Transformed Rules)

Evaluation of the correctly identified direction of German sound changes. Feature pairs with correct direction are marked in italics. Column "C" stands for the most relevant condition; "AA" for *alveolar_affricate*. For the different methods, see Sect. 4.1.1.4.

Pair	Phonological Rule	C	Phonetic Plausibility	Articulatory Closer to the Sound Environment	Phonetic Distribution
1 vibrant-central vibrant-near_open	ʁ > ɐ / _$	ˈiː_ ə	vibrant > central / near_open	vibrant > central / near_open	central / near_open > vibrant
2 mid-near_open	R > R̥ / ə>0_	_ʀ	-	-	near_open > mid
5 voiceless-voiced	D > T / _#	ˈeː_ t	-	voiced > voiceless	voiced > voiceless
9 front-back near_close-open_mid	*umlaut*	_ç	-	back > front	back > front
11 centralized-open_mid	*e > *i / _NC, C_0i,j,u	AA	-	-	centralized > near_close
13 χ-ç	x > ç / {-V[-front]}_	_#	-	-	fricative > palatal
14 short-long	V > [-long] / _[-stress]	AA	-	-	long > short
21 long-velar	ŋ > 0 / V>V:_h,x	ʁ_ŋ	-	long > velar	velar > long
34 ʀ-ᴢ	*z > *ɹ̥ / _	-	ʀ > ᴢ	-	ᴢ > ʀ
TOTAL			2	3	4
			0.222222	0.33333	0.444444

App B.19 Conditions of the Phonological Rules (Paradigmatic, Transformed)

Evaluation of the identified conditions of German phonological rules. If the direction was determined incorrectly, the condition was adjusted (column "Adj. Rule"). For more details, see Sect. 7.5.1.4. Legend: P = number of identified conditions (threshold > 0.9, see Sect. 8.1.3 and 8.2.4), TP = true positives.

Pair	Phonological Rule	Adj. Rule	P	TP	Precision	Doublets
1 vibrant-central, vibrant-near_open	r,ʁ,ʀ > ɐ / _C_0	_V	66	57	0.863636	0.982456
2 mid-near_open	/ə/ > Ø / _C[+R]C_0	_/r ʁ,ʀ/	21	15	0.714286	0.866667
5 voiceless-voiced	[-R]>[-voice] / _[-R]	-	109	42	0.385321	0.928571
13 χ-ç	ç > [-front] / V[-front]_	_V[-front]	26	9	0.346154	0.888889
14 short-long	V > [-long] / _[-stress]	-	17	0	0	-
TOTAL			239	123	0.514644	0.916646

App B.20 Conditions of the Sound Changes (Paradigmatic, Transformed Rules)

Evaluation of the identified conditions of German sound changes. If the direction was determined incorrectly, the condition was adjusted (column "Adj. Rule"). For more details, see Sect. 7.5.1.4. Legend: P = number of identified conditions (threshold > 0.9, see Sect. 8.1.3 and 8.2.4), TP = true positives.

Pair	Sound Change	Adj. Rule	P	TP	Precision	Doublets
1 vibrant-central vibrant-near_open	ʁ > ɐ / _$	_V	66	57	0.863636	0.982456

2 mid-near_open	R > R̥ / ə>0_	_/r ʁ,ʀ/	21	15	0.714286	0.866667
5 voiceless-voiced	D > T / _#	-	109	42	0.385321	0.928571
9 front-back near_close-open_mid	*umlaut*	-	21	0	0	
11centralized-o-pen_mid	*e > *i / _NC, C_0i,j,u	-	6	1	0.166667	0
13 χ-ç	x > ç / {-V[-front]}_	_V[-front]	26	9	0.346154	0.888889
14 short-long	V > [-long] / _[-stress]	-	17	0	0	
21 long-velar	ŋ > 0 / V>V:_h,x	-	23	0	0	
34 ʀ-z	*z > *ɹ̝ / _	-	0	0	-	
TOTAL			185	146	0.789189	0.727766
only diachronic sound changes			32	10	0.3125	

App B.21 Relevance Measures for Sound Correspondences

Comparison of relevance measures for paradigmatic sound correspondences in German:

(1) Frequency: absolute number of lexemes with the corresponding alternation.

(2) $Jaccard(x, y) = \frac{freq(x,y)}{freq(x)+freq(y)-freq(x,y)}$

(3) $weightFreq(x, y) = \frac{freqAllo(x,y)}{Levenshtein(x,y)+1}$ *(frequency weighted by the Levenshtein distance)*

(4) Jaccard index (as 2) used with IPA-transcribed data (the different variants of consonantal /r/ result from inconsistent transcriptions in Wiktionary).

Frequency	Jaccard(x,y)	weightFreq(x,y)	Jaccard(x,y) IPA
ä-a (4611.0)	ä-a (0.08077)	ü-u (18.91358)	ˈœː-œː (0.50000)
ü-u (1532.0)	ö-o (0.04231)	ü-a (16.37500)	ˈæ-æ (0.22222)
ö-o (1339.0)	ü-u (0.04012)	ä-u (14.28571)	ʀ-ɐ̯ (0.13305)
i-e (379.0)	ä-i (0.00476)	ü-ö (12.00000)	l̥-l (0.09313)
o-i (374.0)	o-i (0.00468)	z-c (8.00000)	ˈuː-uː (0.07744)
ä-e (316.0)	ß-s (0.00421)	ü-o (5.90000)	ˈyː-yː (0.06718)
ä-i (299.0)	ä-o (0.00402)	ä-a (5.17508)	ʁ-ɐ̯ (0.06608)

App B.22 Determination of Conditions (Morphophonemic Method)

On the left of the following table, the percentage distribution of environments of the German allophone [ç] in alternating paradigms is given (see Sect. 8.1.3.1). Values close to 1.0 reflect the conditions for [ç], values close to 0.0 the conditions for [x]. The values of the difference measure, which determines the relevance of conditions, are in the right column. High positive values indicate relevant conditions of [ç], negative values those of [x]. Correct environments of [ç] are marked with (C), the true environments of [x] with (X), and neutral environments with (N).

Percentage Distribution		Difference Measure	
0.0	C: ˈɪ, ˈy X: ˈɔ_t͡s, ˈaʊ̯_, ˈa_t, ˈɔ, ˈaː_ə, a_, ˈaʊ̯_f, ˈaʊ̯_p, ˈaː_, ˈaʊ̯_ɪ, a_#, a_t͡s, a, ˈɔ_n̩, ˈɔ_, ˈa_, ˈa_b, ˈa_t͡s N: ʔ, _f, _ɪ, _p, x,	-0.043130	N: ə
0.2	N: _t͡s	-0.035714	X: ˈɔ, ˈɔ_, ˈa_
0.25	X: ˈaʊ̯ N: _ə,	-0.032689	N: t

0.333333	N: l̩, _b	-0.029762	X: ˈɔ_n̩
0.363636	X: ˈa	-0.026347	X: ˈa
0.388889	N: ə	-0.025078	X: ˈaʊ̯
0.4	X: ˈaː N:g	-0.023810	X: ˈa_t
0.454545	N: t͡s	-0.018052	N: t͡s
0.5	C:ˈœ, ˈɛ_l, ˈɛ_t͡s, ˈɔɪ̯_ X: ʊ, aː, N: ʃ,	-0.017857	X: ˈaʊ̯_
	C: ʁ_b, ˈœ_l̩, ˈœ_, ˈiː, ˈɔɪ̯_l, ˈeː	-0.015515	N: _t͡s
	C: ˈɔɪ̯, ˈɛ_ə, ʁ_ X: u N: _t, _l̩, z, f, k,	-0.012100	N: _t, k
0.538462	N: t	-0.011905	X: ˈaː_ə, a_, ˈaː_, a
0.571429	X: ˈoː	-0.009563	N: _ə
0.571429	C: ˈaɪ̯	-0.007221	X: ˈaː, g
0.6	N: p	-0.005952	C: ˈyː, ˈɪ, X: ˈɔ_t͡s, ˈaʊ̯_f, ˈaʊ̯_p, ˈaʊ̯_ɪ, a_#, a_t͡s, ˈa_b, ˈa_t͡s N: _f, _ɪ, _p, ʔ,
0.666667	C: ˈøː_s, ˈøː_ N: _s, _l, m, ɐ, ŋ		
0.7	N: ç, s	-0.004879	X: f, z
0.702128	N: b	-0.003610	X: l̩, _b
0.714286	N: h, _#, ʁ	-0.002537	C: ˈaɪ̯ X: ˈoː
0.75	C: ˈøː, ɪ, X: ˈʊ N: d, v	-0.000098	N: b
0.761905	N: n	0.001073	N: p

0.763158	N: n̩	0.002342	C: ˈœ, ˈɛ_l, ˈɛ_t͡s, ˈɔɪ̯_, ʁ_b, ˈœ_l̩, X: ʊ, aː, N: ʃ,
0.769231	N: l		C: ˈœ_, ˈiː, ˈɔɪ̯_l, ˈeː, ˈɔɪ̯, ˈɛ_ə, ʁ_ X: u N: _l̩
0.777778	N: ɐ̯	0.003415	N: ɐ
0.8	C: ɛ	0.004489	N: ç, s
0.823529	N: _n̩	0.004684	C: ˈøː_s, ˈøː_ N: ŋ, m, _l, _s,
0.833333	C: ɪ_,ɪ_#,ˈɛ_t	0.005172	N: ʁ
0.875	N: ʀ	0.005757	N: _#, h
0.965517	C: ˈɛ_n̩	0.007026	C: ˈøː X: ˈʊ N: d, v
0.972972	C: ˈɛ_	0.009173	C: ɪ
0.973684	C: ˈɛ	0.010441	N: ɐ̯
		0.011514	N: l
		0.011710	C: ˈɛ_t, ɪ_#, ɪ_
		0.012783	C: ɛ
		0.013661	N: n
		0.020297	N: n̩
		0.026835	N: ʀ
		0.035812	N: _n̩
		0.065574	C: ˈɛ_n̩
		0.084309	C: ˈɛ_
		0.086651	C: ˈɛ

App B.23 Sample for the Significance Test of Sound Environments

The following table lists the probability values of the Chi-squared four-fold test for each environment of [ç] (see Sect. 8.1.3.2). The degree of freedom is 1.

Φ (χ²)		Φ (χ²)	
1.0	ˈɛ, _n̩, ˈɛ_, b, ˈɛ_n̩	0.680144	k
0.999999	ˈɛ_t	0.653983	h
0.999999	n̩	0.633468	l
0.999935	ʁ	0.597130	ɛ
0.941335	t	0.328789	ɐ̯
0.935807	ʀ	0.175638	ɪ
0.907222	ə	0.141340	s
0.810922	_t	0.038314	t͡s
0.801561	n	-	others

App B.25 Preliminary Result of the Transformed Paradigmatic Method

Preliminary result of the twelve most relevant sound changes (without specification of conditions) before reducing doublets.

	Pair	Corresponding Sound Correspondences	Relevance
1	ʀ-ɐ̯	ʀ-ɐ̯	0.779818
2	trill-central	ʀ-ɐ̯, ʀ-ɐ	0.755414
3	trill-near_open	ʀ-ɐ̯, ʀ-ɐ	0.755414
4	uvular-near_open	ʀ-ɐ̯, ʀ-ɐ, ʁ-ɐ̯, ʁ-ɐ	0.743099
5	trill-semivowel	ʀ-ɐ̯, ʀ-aɪ̯	0.739794
6	uvular-semivowel	ʀ-ɐ̯, ʀ-aɪ̯, ʁ-aɪ̯, ʁ-ɐ̯	0.731047

7	uvular-central	ʀ-ɐ̯, ʀ-ɐ, ʁ-ɐ, ʁ-ɐ̯, ʁ-ə	0.725274
8	voiced-near_open	ʀ-ɐ̯, ʀ-ɐ, ʁ-ɐ, ʁ-ɐ̯, z-ɐ̯, l-ɐ̯	0.708965
9	ʁ-ɐ̯	ʁ-ɐ̯	0.702304
10	central-voiced	ʁ-ɐ̯, ʀ-ɐ, ʁ-ɐ, ə-ŋ, ʀ-ɐ̯, ə-m̩, ə-l̩, ə-n̩, ə-ŋ, z-ɐ̯, l-ɐ̯, ʁ-ə	0.687109
11	voiced-semivowel	ʀ-ɐ̯, z-aɪ̯, ʁ-aɪ̯, ʀ-aɪ̯, b-aɪ̯, ʁ-ɐ̯, ˈaɪ̯-n, ˈɔɪ̯-ŋ, ɐ̯-z, ɐ̯-l,ˈaʊ̯-v n-aɪ̯	0.665345
12	mid_vowel-near_open	ə-ɐ	0.653292

App B.26 Determination of the Threshold of *condMeasure*

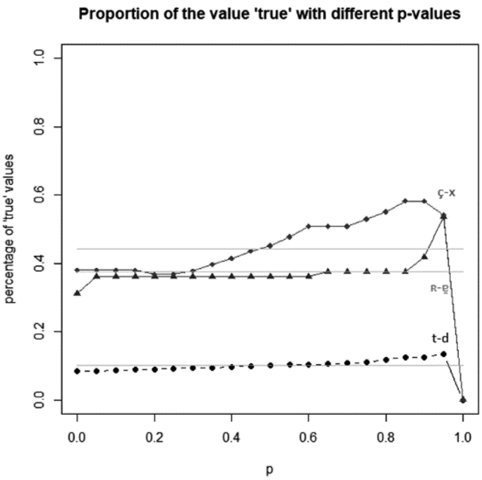

Figure B.26: On The following figure shows the percentage of *condMeasure* values (see Sect. 8.1.3.2) with true conditions calculated for different threshold values *p*. The average values are marked with grey vertical lines.

Appendix C: Derivational and Semantic Approach

App C.1 Result Derivational Approach (Single Sounds)

Result list of the morphophonemic method (derivational approach, single sounds) containing the 100 sound pairs with the highest Jaccard indices and the assignment to a phonological rule (PR), a Pre-New High German (NHG), a Pre-Middle High German (MHG), or a Pre-Old High German (OHG) sound change. Legend: PR = phonological rules, SC = sound change, FV = free variant, Abl = *ablauting* alternation, TS = transcription error, WC = wrong correspondence, R = sonorant, Sup = suppletive forms, I = i, ī, j, iu; →*NUMBER* = doublet of pair *NUMBER*.

	Pair	Jaccard		PR	SC	Example
1	ʁ|ɐ̯	0.671928	PR/ SC	r / ʀ / ʁ > ɐ̯ / _C_0	ʁ > ɐ / _$ (NHG)	
2	ʀ|ɐ̯	0.665096	FV			
3	ʁ|ʀ	0.664056	FV			
4	ˈaɪ̯|aɪ̯	0.633627	TS			
5	χ|x	0.624219	FV			
6	ˈɛ|ˈa	0.600168	SC/ Abl		ˈa > [ˈæ] / _...I (MHG)	
7	ˈɛː|ˈaː	0.599594	SC		ˈā > ˈæ [ɛː] / _...I (MHG)	
8	ˈɛ|ɛ	0.588867	TS			
9	ˈa|a	0.581887	TS			
10	ˈɔ|ɔ	0.581556	TS			
11	t|d	0.580951	PR/ SC	d > t / _[-R]$_0$.	d > t / _# (NHG)	

12	ˈʊ|ʊ	0.580426	TS			
13	g|k	0.576189	PR/ SC	g > k / _$[-R]_0$.	g > k / _# (NHG)	
14	ˈɪ|ɪ	0.560090	TS			
15	ˈɔɪ̯|ˈaʊ̯	0.553653	SC		ˈou > ˈöu / _...I (MHG)	
16	ˈaʊ̯|aʊ̯	0.543424	TS			
17	ə|n̩	0.542889	PR/ SC	ə > 0 / _$C[+R]C_0$.	n > n̥ / ə>0_ (NHG)	
18	p|b	0.542082	PR/ SC	b > p / _$[-R]_0$.	b > p / _# (NHG)	
19	ɪ|n̩	0.529686	-			
20	ˈaː|aː	0.523985	TS			
21	ˈɔ|ˈœ	0.522326	SC		ˈo > ˈö / _...I (MHG)	
22	ˈoː|o	0.518661	PR	oː > o / _[-stress]		*Kategorie: kategorisch*
23	z|s	0.515173	PR/ SC	z > s / _$[-R]_0$.	z > s / _# (NHG)	*grasen:Gras*
24	ˈʏ|ˈʊ	0.511040	SC		ˈu > ˈü / __C {-lt,ld,-NC} ...I (MHG)	
25	ˈaː|a	0.500366	TS			
26	ˈɔɪ̯|ɔɪ̯	0.498262	TS			
27	ɪ|ə	0.497179	-			
28	ˈeː|eː	0.493869	TS			
29	t͡s|t	0.492446	MA / SC?			*Abstraktion : abstrakt*
30	ə|ɐ	0.489651	PR/ SC	ə > 0 / _$C[+R]C_0$.	ə > 0 / _ɐ (NHG)	

31	χ|ç	0.487952	PR	ç > x / V[-front]_	x > ç; γ > ʝ / {-V[-front]}_ (NHG)	
32	ˈiː|ˈaɪ̯	0.484526	MA			
33	g|ç	0.480926	PR	g > ç / ɪ[-stress]_C_0.		
34	ˈɪ|ˈaɪ̯	0.473446	MA			
35	ə|l	0.465947	-			
36	ˈyː|ˈuː	0.462034	SC		ˈū > [ˈyː] / _...I (MHG)	
37	ə|l̥	0.458465	PR/ SC	l > l̥ / .C_0_C_0. and ə > 0 / _C[+R]C0.	l > l̥ / ə>0_ (NHG)	
38	ˈɛː|aː	0.451422	SC/ MA			*Freitag :* *freitäglich* (→7)
39	ˈʏ|ʏ	0.450291	TS			
40	ˈiː|iː	0.448689	TS			
41	ˈi̯|i̯	0.444206	-			
42	ˈoː|oː	0.443366	TS			
43	ˈɛ|ˈɔ	0.431795	MA			
44	ˈeː|e	0.423326	PR	eː > e / _[-stress]		*nebulös:Ne-* *bel*
45	ˈyː|yː	0.422406	TS			
46	ˈɪ|ˈɛ	0.422061	SC/ Abl		*e > *i / _NC, C_0i,j,u (OHG)	
47	ˈiː|ɪ	0.420911	-			*Demokratie :* *demokratisch*
48	ˈiː|i	0.418336	PR	iː > i / _[-stress]		*politisieren :* *politisch*
49	z|b	0.418292	-			

50	n̩|l	0.409891	-			
51	ˈi|i	0.407288	TS			
52	ˈʊ|ˈɪ	0.405072	MA			
53	ə|t	0.402215	-			
54	ˈaː|ˈa	0.400903	TS/ PR	aː > a / _[-stress]		*Schlag:Schlacht*
55	ˈoː|ˈiː	0.400476	MA			
56	t͡s|d	0.400255	MA / PR?			*Millarde : millardstel*
57	t|n	0.399160	-			
58	ˈiː|ə	0.397440	-			
59	oː|o	0.395574	TS			*Alkohol : Alkoholiker*
60	n̩|l̩	0.395267	-			
61	z|t	0.390758	-			
62	ˈʏ|ʊ	0.388453	MA			→24
63	ʀ|b	0.386744	-			
64	ˈɛ|a	0.382093	SC			→6
65	ˈa|ʀ	0.382019	-			
66	ˈuː|u	0.376746	PR	uː > u / _[-stress]		*Kultur:kultu-rell*
67	ʃ|k	0.376255	-			
68	ɪ|l	0.374483	-			
69	ˈeː|ˈa	0.374018	MA			
70	ʊ|u	0.373504	TS			
71	ˈe|e	0.373033	TS			
72	v|f	0.372947	PR/ SC	v > f / _[-R]$_0$.	v > f / _# (NHG)	
73	z|d	0.372922	-			

74	ˈaʊ̯|ʀ	0.372530	-			
75	n|f	0.372347	-			
76	f|b	0.372064	-			
77	z|l	0.371495	-			
78	ˈiː|ˈaː	0.369144	MA			
79	ˈeː|ˈaɪ̯	0.368341	-			
80	ʀ|l	0.367222	-			
81	ʀ|ɐ	0.365995	PR	r, ʀ, ʁ > ɐ / .C_0_C_0.		
82	ç|x	0.365951	FV			
83	ɪ|t	0.365531	-			
84	ɪ|z	0.364842	-			
85	ˈuː|uː	0.364148	TS			
86	ʃ|z	0.364114	-			
87	ˈaɪ̯|ɪ	0.363705	MA			
88	ɪ|i	0.362657	TS			
89	ɪ|b	0.359135	-			
90	ʀ|t	0.358826	-			
91	ˈa|b	0.357939	-			
92	ʃ|f	0.353930	-			
93	ç|k	0.353544	TS/ SC		x > k / _s (NHG)	*brechen:Bro-cken*
94	f|d	0.353242	-			
95	ˈiː|ˈa	0.352531	MA			
96	ˈa|ʃ	0.352159	-			
97	ʁ|t	0.349345	-			
98	l|b	0.348519	-			
99	ˈa|ɛ	0.346726	SC? / TS			*anfangen : anfänglich* (→6)
100	ˈa|f	0.345038	-			

App C.2 Phonological Rules (Derivational, Single Sounds)

correctly identified	17	gold standard	26
free variants and doublets	8		
wrongly identified	75		
TOTAL	100		

Evaluation for the phonological rules using different threshold values. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	Correct Sound Pair	Number of Identified Pairs	Match Precision	Match Recall	Match F-Score
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-
0.8	0	0	-	-	-
0.75	0	0	-	-	-
0.7	0	0	-	-	-
0.65	1	1	1	0.038462	0.074074
0.6	0	2	0.333333	0.038462	0.068966
0.55	2	9	0.25	0.115385	0.157895
0.5	4	10	0.318182	0.269231	0.291667
0.45	*4*	*13*	*0.314286*	*0.423077*	*0.360656*
0.4	3	17	0.269231	0.538462	0.358974
0.35	3	37	0.191011	0.68	0.298246
0.3	0	3	0.184783	0.653846	0.288136
0.25	-	-	0.184783	0.653846	0.288136

0.2	-	-	0.184783	0.653846	0.288136
0.15	-	-	0.184783	0.653846	0.288136
0.1	-	-	0.184783	0.653846	0.288136
0.0	-	-	0.184783	0.653846	0.288136
TOTAL	17	92			

App C.3 Old High German Sound Changes (Derivational, Single Sounds)

Since Old High German			
correctly identified	18	gold standard	216
free variants and doublets	15	(including 7 PR)	
wrongly identified	67		
TOTAL	100		
Only Old High German			
correctly identified	1	gold standard	104
free variants and doublets	15		
wrongly identified	67		
TOTAL	100		

Evaluation for the sound changes since (Pre-)Old High German times using different threshold values. All sound changes listed in App. A.1 are considered. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	Correct Sound Pair	Number of Identified Pairs	Match Precision	Match Recall	Match F-Score
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-
0.8	0	0	-	-	-

0.75	0	0	-	-	-
0.7	0	0	-	-	-
0.65	1	1	1	0.004630	0.009217
0.6	1	2	0.666667	0.009259	0.018265
0.55	4	9	0.5	0.027778	0.052632
0.5	5	9	0.523810	0.050926	0.092827
0.45	4	12	0.454545	0.069444	0.120482
0.4	*1*	*14*	*0.340426*	*0.074074*	*0.121673*
0.35	2	35	0.219512	0.083333	0.120805
0.3	0	3	0.211765	0.083333	0.119601
0.25	-	-	0.211765	0.083333	0.119601
0.2	-	-	0.211765	0.083333	0.119601
0.15	-	-	0.211765	0.083333	0.119601
0.1	-	-	0.211765	0.083333	0.119601
0.0	-	-	0.211765	0.083333	0.119601
TOTAL	18	85			

App C.4 Middle High German Sound Changes (Derivational, Single Sounds)

Since Middle High German			
correctly identified	17	gold standard	112
free variants and doublets	15	(including 7 PR)	
wrongly identified	68		
TOTAL	100		
Only Middle High German			
correctly identified	11	gold standard	37
free variants and doublets	15		
wrongly identified	74		
TOTAL	100		

Evaluation for the sound changes since (Pre-)Middle High German times using different threshold values. All Pre-Middle and Pre-New High German sound changes listed in App. A.1 are considered. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	**Correct Sound Pair**	**Number of Identified Pairs**	**Match Precision**	**Match Recall**	**Match F-Score**
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-
0.8	0	0	-	-	-
0.75	0	0	-	-	-
0.7	0	0	-	-	-
0.65	1	1	1	0.008929	0.017699
0.6	1	2	0.666667	0.017857	0.034783
0.55	4	9	0.5	0.053571	0.096774
0.5	5	9	0.523810	0.098214	0.165414
0.45	*4*	*12*	*0.454545*	*0.133929*	*0.206897*
0.4	0	14	0.319149	0.133929	0.188679
0.35	2	35	0.207317	0.151786	0.175258
0.3	0	3	0.2	0.151786	0.172589
0.25	-	-	0.2	0.151786	0.172589
0.2	-	-	0.2	0.151786	0.172589
0.15	-	-	0.2	0.151786	0.172589
0.1	-	-	0.2	0.151786	0.172589
0.0	-	-	0.2	0.151786	0.172589
TOTAL	17	85			

App C.5 New High German Sound Changes (Derivational, Single Sounds)

Since New High German			
correctly identified	6	gold standard	74
free variants and doublets	15	(including 7 PR)	
wrongly identified	79		
TOTAL	100		

Evaluation for the sound changes since (Pre-)New High German times using different threshold values. All (Pre)-New High German sound changes listed in App. A.1 are considered. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	Correct Sound Pair	Number of Identified Pairs	Match Precision	Match Recall	Match F-Score
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-
0.8	0	0	-	-	-
0.75	0	0	-	-	-
0.7	0	0	-	-	-
0.65	1	1	1	0.013514	0.026667
0.6	0	2	0.333333	0.013514	0.025974
0.55	0	9	0.083333	0.013514	0.023256
0.5	1	9	0.095238	0.027027	0.042105
0.45	*3*	*12*	*0.151515*	*0.067568*	*0.093458*
0.4	0	14	0.106383	0.067568	0.082645
0.35	1	35	0.073171	0.081081	0.076923

0.3	0	3	0.070588	0.081081	0.075472
0.25	-	-	0.070588	0.081081	0.075472
0.2	-	-	0.070588	0.081081	0.075472
0.15	-	-	0.070588	0.081081	0.075472
0.1	-	-	0.070588	0.081081	0.075472
0.0	-	-	0.070588	0.081081	0.075472
TOTAL	6	85			

App C.6 Correct Complementary Sounds (Derivational, Single Sounds)

Evaluation of the correctly identified complementary sounds. Sound pairs whose sounds belong to a German phonological rule or sound change are marked in italics. Legend: R = any sonorant; I = i,ī,j,iu.

<table>
<tr><th colspan="3">Phonological Rules</th><th colspan="3">Sound Changes</th></tr>
<tr><th></th><th>Pair</th><th>Rule</th><th></th><th>Pair</th><th>Rule</th></tr>
<tr><td>1</td><td>ʁ|ɐ̯</td><td>r / ʀ / ʁ > ɐ̯ / _C_0</td><td>1</td><td>ʁ|ɐ̯</td><td>ʁ > ɐ / _$</td></tr>
<tr><td>11</td><td>t|d</td><td>d > t / _[-R]$_0$.</td><td>6</td><td>ˈɛ|ˈa</td><td>ˈa > [ˈæ] / _...I</td></tr>
<tr><td>13</td><td>g|k</td><td>g > k / _[-R]$_0$.</td><td>7</td><td>ˈɛː|ˈaː</td><td>ˈā > ˈæ [ˈɛː] / _...I</td></tr>
<tr><td>17</td><td>ə|n̩</td><td>n > n̥ / ə>0_ or ə > 0 / _C[+R]C_0.</td><td>11</td><td>t|d</td><td>d > t / _#</td></tr>
<tr><td>18</td><td>p|b</td><td>b > p / _[-R]$_0$.</td><td>13</td><td>g|k</td><td>g > k / _#</td></tr>
<tr><td>22</td><td>ˈoː|o</td><td>oː > o / _[-stress]</td><td>15</td><td>ˈɔɪ̯|ˈaʊ̯</td><td>ˈou > ˈöu [ˈœʏˀ] / _...I</td></tr>
<tr><td>23</td><td>z|s</td><td>z > s / _[-R]$_0$.</td><td>17</td><td>ə|n̩</td><td>n > n̥ / ə>0_ or ə > 0 / _C[+R]C_0.</td></tr>
<tr><td>30</td><td>ə|ɐ</td><td>ə > 0 / _C[+R]C_0.</td><td>18</td><td>p|b</td><td>b > p / _#</td></tr>
<tr><td>33</td><td>g|ç</td><td>g > ç / ɪ[-stress]_C_0.</td><td>21</td><td>ˈɔ|ˈœ</td><td>ˈo > ˈö [ˈœˀ] / _...I</td></tr>
<tr><td>37</td><td>ə|l̩</td><td>l > l̥ / .C_0_C_0. and ə > 0 / _C[+R]C_0.</td><td>23</td><td>z|s</td><td>z > s / _#</td></tr>
</table>

44	*ˈeː|e*	*eː > e / _[-stress]*	*24*	*ˈʏ| ˈʊ*	*ˈu > ˈü [ʏ$^{?}$] / _C{ -lt,ld,-NC}...I*
48	*ˈiː|i*	*iː > i / _[-stress]*	30	ə|ɐ	ə > 0 / _ɐ
54	*ˈaː| ˈa*	*aː > a / _[-stress]*	*36*	*ˈyː| ˈuː*	*ˈū > ˈiu [ˈyː] / _...I*
66	*ˈuː|u*	*uː > u / _[-stress]*	37	ə|l̩	l > l̥ / ə>0_
72	*v|f*	*v > f / _[-R]$_0$.*	46	ˈɪ|ˈɛ	*e > *i / _NC, C$_0$i,j,u
81	ʀ|ɐ	r / ʀ / ʁ > ɐ / .C$_0$_C$_0$.	*72*	*v|f*	*v > f / _#*
82	*ç|x*	*ç > x / V[-front]_*	*82*	*ç|x*	*x > ç; γ > j / {-V[-front]}_*
			93	ç|k	x > k / _s
TOTAL		13 (0.764706)	TOTAL		10 (0.555556)

App C.7 Direction of the Phonological Rules (Derivational, Single Sounds)

Evaluation of the correctly identified direction of German phonological rules. Sound pairs with correct direction are marked in italics. For the different methods, see Sect. 4.1.1.4. Legend: "C" stands = most relevant condition; R = sonorant.

Pair	Phonological Rule	C	Phonetic Plausibility	Articulatory Closer to the Sound Environment	Phonetic Distribution
1 ʁ|ɐ̯	r / ʀ / ʁ > ɐ̯ / _C$_0$	_ə	*ʁ>ɐ̯*	*ʁ>ɐ̯*	ɐ̯>ʁ
11 t|d	d > t / _[-R]$_0$.	_#	-	*d>t*	*d>t*
13 g|k	g > k / _[-R]$_0$.	_n̩	-	k>g	k>g
17 ə|n̩	n > n̥ / ə>0_ or ə > 0 / _C[+R]C$_0$.	_n	-	ə>n̩	n̩>ə
18 p|b	b > p / _[-R]$_0$.	_t	-	*b>p*	*b>p*
22 ˈoː|o	oː > o / _[-stress]	_ɐ̯	-	-	o>ˈoː

<table>
<tr><td>23 z|s</td><td>z > s / _[-R]$_0$.</td><td>_ə</td><td>-</td><td>s>z</td><td>s>z</td></tr>
<tr><td>30 ə|ɐ</td><td>ə > 0 / _C[+R]C_0.</td><td>_ʁ</td><td>-</td><td>-</td><td>ɐ>ə</td></tr>
<tr><td>33 g|ç</td><td>g > ç / ɪ[-stress]_C_0.</td><td>_n̩</td><td>-</td><td>g>ç</td><td>ç>g</td></tr>
<tr><td>37 ə|l̩</td><td>l > l̥ / .C_0_C_0. and ə > 0 / _C[+R]C_0.</td><td>_l</td><td>-</td><td>ə>l̩</td><td>l̩>ə</td></tr>
<tr><td>44 ˈeː|e</td><td>eː > e / _[-stress]</td><td>l̩</td><td>-</td><td>-</td><td>e>ˈeː</td></tr>
<tr><td>48 ˈiː|i</td><td>iː > i / _[-stress]</td><td>-</td><td>-</td><td>-</td><td>i>ˈiː</td></tr>
<tr><td>54 ˈaː|ˈa</td><td>aː > a / _[-stress]</td><td>ɐ̯</td><td>-</td><td>-</td><td>ˈa>ˈaː</td></tr>
<tr><td>66 ˈuː|u</td><td>uː > u / _[-stress]</td><td>-</td><td>-</td><td>-</td><td>u>ˈuː</td></tr>
<tr><td>72 v|f</td><td>v > f / _[-R]$_0$.</td><td>-</td><td>-</td><td>-</td><td>f>v</td></tr>
<tr><td>81 ʀ|ɐ</td><td>r / ʀ / ʁ > ɐ / .C_0_C_0.</td><td>_ɪ</td><td>-</td><td>ʀ>ɐ</td><td>ɐ>ʀ</td></tr>
<tr><td>82 ç|x</td><td>ç > x / V[-front]_</td><td>ˈɛ</td><td>-</td><td>x>ç</td><td>x>ç</td></tr>
<tr><td colspan="3">TOTAL</td><td>1</td><td>5</td><td>2</td></tr>
<tr><td colspan="3"></td><td>0.083333</td><td>0.416667</td><td>0.166667</td></tr>
</table>

App C.8 Direction of the Sound Changes (Derivational, Single Sounds)

Evaluation of the correctly identified direction of German sound changes. Sound pairs with correct direction are marked in italics. For the different methods, see Sect. 4.1.1.4. Legend: C = most relevant condition; R = sonorant; I = i,ī,j,iu.

Pair	Phonological Rule	C	Phonetic Plausibility	Articulatory Closer to the Sound Environment	Phonetic Distribution
1	ʁ > ɐ / _$	_ə	*ʁ>ɐ̯*	*ʁ>ɐ̯*	ɐ̯>ʁ
6	ˈa > [ˈæ] / _...I	_ç	-	*ˈa>ˈɛ*	*ˈa>ˈɛ*
7	ˈā > ˈæ [ˈɛː] / _...I	-	-	-	*ˈaː>ˈɛː*

11	d > t / _#	_#	-	*d>t*	*d>t*
13	g > k / _#	_n̩	-	k>g	k>g
15	ˈou>ˈöu [ˈœʏˀ] / _...I	-	-	-	*ˈaʊ̯> ˈɔɪ̯*
17	n > n̥ / ə>0_ *or* ə > 0 / _C[+R]C_0.	_n	-	ə>n̩	n̩>ə
18	b > p / _#	_t	-	*b>p*	*b>p*
21	ˈo > ˈö [ˈœˀ] / _...I	n̩	-	-	ˈœ>ˈɔ
23	z > s / _#	_ə	-	s>z	s>z
24	ˈu > ˈü [ˈʏˀ] / _C{-lt,ld,-NC}...I	-	-	-	*ˈʊ> ˈʏ*
30	ə > 0 / _ɐ	_ʁ	-	-	ɐ>ə
36	ˈū > ˈiu [ˈyː] / _...I	-	-	-	*ˈuː> ˈyː*
37	l > l̥ / ə>0_	_l	-	ə>l̩	l̩>ə
46	*e > *i / _NC, C_0i,j,u	-	-	*ˈɛ> ˈɪ*	*ˈɛ> ˈɪ*
72	v > f / _#	-	-	-	f>v
82	x > ç; ɣ > ʝ / {-V[-front]}_	ˈɛ	-	*x>ç*	*x>ç*
93	x > k / _s	ˈɛ_n̩	-	k>ç	k>ç
TOTAL			1	6	9
			0.083333	0.5	0.75

App C.9 Conditions of the Phonological rules (Derivational, Single Sounds)

Evaluation of the identified conditions of German phonological rules. If the direction was determined incorrectly, the condition was adjusted (column "Adjusted rule"). For more details, see Sect. 7.5.1.4. Legend: Positives = number of identified conditions (threshold > 0.9, see Sect. 8.1.3 and 8.2.4), TP = true positives. *R* stands for any sonorant.

Evaluation of the identified conditions of German phonological rules. If the direction was determined incorrectly, the condition was adjusted (column “Adj. Rule”). For more details, see Sect. 7.5.1.4. Legend: P = number of identified conditions (threshold > 0.9, see Sect. 8.1.3 and 8.2.4), TP = true positives, R = sonorant.

Pair	**Phonological Rule**	**Adj. Rule**	**P**	**TP**	**Precision**	**Doublets**
1 ʁ|ɐ̯	r / ʀ / ʁ > ɐ̯ / _C_0	_V	15	6	0.4	0.833333
11 t|d	d > t / _[-R]$_0$.		13	5	0.384615	0.8
13 g|k	g > k / _[-R]$_0$.	_V, _C [+voiced]	8	8	1	0.75
17 ə|n̩	n > n̥ / ə>0_ *or* ə > 0 / _C[+R]C_0.	_/n/	2	2	1	0.5
18 p|b	b > p / _[-R]$_0$.		12	6	0.5	0.666667
22 ˈoː|o	oː > o / _[-stress]		2	0	0	-
23 z|s	z > s / _[-R]$_0$.		4	4	1	0.75
30 ə|ɐ	ə > 0 / _C[+R]C_0.	_r,ʁ,ʀ	1	1	1	0
33 g|ç	g > ç / ɪ[-stress]_C_0.	_V	4	1	0.25	0
37 ə|l̩	l > l̥ / .C_0_C_0. *and* ə > 0 / _C[+R]C_0.	_l	1	1	1	0
44 ˈeː|e	eː > e / _[-stress]		1	0	0	-
48 ˈiː|i	iː > i / _[-stress]		0	0	-	-
54 ˈaː|ˈa	aː > a / _[-stress]		11	0	0	-
66 ˈuː|u	uː > u / _[-stress]		0	0	-	-
72 v|f	v > f / _[-R]$_0$.		0	0	-	-
81 ʀ|ɐ	r / ʀ / ʁ > ɐ / .C_0_C_0.	_V	1	1	1	0
82 ç|x	ç > x / V[-front]_	V[+front]_ , C_	3	3	1	0.666667
TOTAL			78	38	0.487179	0.451515

App C.10 Conditions of the Sound Changes (Derivational, Single Sounds)

Evaluation of the identified conditions of German sound changes. If the direction was determined incorrectly, the condition was adjusted (column “Adj. Rule”). For more details, see Sect. 7.5.1.4. Legend: P = number of identified conditions (threshold > 0.9, see Sect. 8.1.3 and 8.2.4), TP = true positives, R = sonorant, I = i,ī,j,iu..

Pair	Sound Change	Adj. Rule	P	TP	Precision	Doublets
1 ʁ|ɐ̯	ʁ > ɐ / _$	_V	15	6	0.4	0.833333
6 ˈɛ|ˈa	ˈa > [ˈæ] / _...I		1	0	0	-
7 ˈɛː|ˈaː	ˈā > ˈæ [ˈɛː] / _...I		0	0	.	-
11 t|d	d > t / _#		13	5	0.384615	0.8
13 g|k	g > k / _#	_V,_C[+voiced]	8	8	1	0.75
15 ˈɔɪ̯|ˈaʊ̯	ˈou > ˈöu [ˈœʏˀ] / _...I		0	0	-	-
17 ə|n̩	n > n̥ / ə>0_ *or* ə > 0 / _C[+R]C_0.	_/n/	2	2	1	0.5
18 p|b	b > p / _#		12	6	0.5	0.666667
21 ˈɔ|ˈœ	ˈo > ˈö [ˈœˀ] / _...I		1	0	0	-
23 z|s	z > s / _#		4	4	1	0.75
24 ˈʏ|ˈʊ	ˈu > ˈü [ˈʏˀ] / _C{-lt,ld-NC}...I		0	0	-	-
30 ə|ɐ	ə > 0 / _ɐ	_r,ʁ,ʀ	1	1	1	0
36 ˈyː|ˈuː	ˈū > ˈiu [ˈyː] / _...I		0	0	-	-
37 ə|l̩	l > l̥ / ə>0_	_l	1	1	1	0
46 ˈɪ|ˈɛ	*e > *i / _NC, C_0i,j,u		0	0	-	-
72 v|f	v > f / _#		0	0	-	-

82 ç|x	x > ç; γ > ʝ / {-V[-front]}_	V[+front]_ , C_	3	3	1	0.666667
93 ç|k	x > k / _s		2	0	0	-
TOTAL			63	36	0.571429	0.551852
only diachronic soun d changes			4	0	0	

App C.11 Result Paradigmatic Approach (Transformed Rules)

Result list of the morphophonemic method (derivational approach, transformed rules) containing the 100 sound pairs with the highest Jaccard indices and the assignment to a phonological rule (PR), a Pre-New High German (NHG), a Pre-Middle High German (MHG), or a Pre-Old High German (OHG) sound change. Legend: PR = phonological rules, SC = sound change, FV = free variant, Abl = *ablauting* alternation, TS = transcription error, WC = wrong correspondence, R = sonorant, Sup = suppletive forms, I = i, ī, j, iu; →*NUMBER* = doublet of pair *NUMBER*.

Pair	Jaccard		Sound Pairs	PR	SC
1 uvular-near_open	0.649413	PR/ SC	ʁ-ɐ̯, ʀ-ɐ̯, ʀ-ɐ, ʁ-ɐ	r,ʁ,ʀ > ɐ / _C_0	ʁ > ɐ / _$ (NHG)
2 stressed-unstres-sed	0.559015	TS	ˈaɪ̯-aɪ̯, ˈɛ-ɛ, ˈa-a, ˈɔ-ɔ, ˈʊ-ʊ, ˈɪ-ɪ, ˈaː-aː, ˈoː-o, ˈiː-ə, …		
3 -	0.553090	FV	ʁ-ʀ		
4 open_mid-open	0.542299	→ 7	ˈɛ-ˈa, ˈɛː-ˈaː, ˈɔɪ̯-ˈaʊ̯, ˈɔɪ̯-aʊ̯,		

			ɛ-a, ˈɛː-ˈa, ˈɔ-ˈa, ˈɛː-ˈaɪ̯, …		
5 mid-syllabic	0.530939	PR/ SC	ə-n̩, ə-l̩, ə-m̩	/ə/ > Ø / _C[+sonorant]C_0	R > R̥ / ə>0_ (NHG)
6 voiceless-voiced	0.517585	PR/ SC	t-d, g-k, p-b, z-s, g-ç, t͡s-d, t-n, z-t, v-f, n-f, f-b, ʃ-z, ʀ-t, …	[-sonorant] > [-voice] / _[-sonorant]	D > T / _# (NHG)
7 back-front	0.498351	SC	ˈɔ-ˈœ, ˈʏ-ˈʊ, ˈyː-ˈuː, ˈøː-ˈoː, ˈɛ-ˈɔ, ˈʊ-ˈɪ, ˈoː-ˈiː, ˈɔ-ˈeː, …		*umlaut* (MHG)
8 short-long	0.483365	PR/ SC	ˈɪ-ˈaɪ̯, ˈaː-ˈa, oː-o, ˈeː-ˈa, ˈiː-ˈa, ˈɪ-ˈeː, ɪ-aɪ̯, aː-	V > [-long] / _[-stress]	V[+stress, -long] > V[+stress, +long] / _${-t,-m,-er}, _r#, _rC[+alveolar] (NHG)

			a, ˈaɪ̯-ˈa, …		
9 -	0.474925	-	ɪ-ə		
10 unstressed-syllabic	0.469934	-	ɪ-n̩, n̩-eː, ɐ-n̩, u-l̩, ɪ-l̩, n̩-aː, n̩-aɪ̯, oː-n̩, ɛ-l̩, ʊ-n̩, ɪ-ŋ̍, l̩-i, ɐ-l̩, ɪ-m̩		
11 close-near_close close-centralized	0.467738	MA /TS	ˈiː-ˈaɪ̯, ʊ-u, ɪ-i, ˈɔɪ̯-ˈiː, ˈɔɪ̯-ˈuː, ˈɪ-ˈiː, ˈuː-ˈaɪ̯, ˈuː-ˈaʊ̯, ˈyː-ˈaɪ̯, ˈɔɪ̯-ˈyː		
12 mid-lateral	0.465947	TS	ə-l		
13 affricate-plosive	0.458750	TS	t͡s-t, t͡s-k, p͡f-p, t͡s-p, t͡ʃ-t, p͡f-k, ʔ-t͡s		
14 -	0.453264	-	ə-ɐ		
15 plosive-fricative	0.441488	SC/ -	z-b, ʃ-k, z-d, ç-k, ʁ-b, ʃ-t, k-f, t-f,		x > k / _s (NHG)

			t-s, k-h, χ-k, v-d, ɡ-v, …		
16 voiceless-vowel	0.435663	-	ə-t, ɪ-t, ˈa-ʃ, ˈa-f, ɪ-t͡s, ˈaɪ̯-t, ə-s, ˈa-h, ə-f, ˈeː-t, ə-t͡s, ˈiː-t, …		
17 nasal-approximant	0.433095	-	n̩-l, n̩-l̩, n-l, m-l, m-j, ŋ-l, ŋ-l̩		
18 vowel-voiced	0.429519	-	ˈa-ʀ, ɪ-l, ˈaʊ̯-ʀ, ɪ-z, ɪ-b, ˈa-b, ˈaɪ̯-n, ɪ-n, ˈɛ-b, ˈaɪ̯-b, …		
19 uvular-alveolar	0.415429	SC	ʀ-l, ʁ-n, ʀ-z, ʀ-n, ʁ-l, ʁ-z, ʀ-d, ʀ-n̩, ʁ-n̩, ʀ-r		*z > *ɹ̠ / _ (OHG)

20 alveolar-bilabial	0.412981	PR	l-b, n-m, d-b, n̩-m̩, n-b, m-d, n̩-b, t-p, z-m, n̩-m	N[+syllabic] > [α place] / [-continuant, -sonorant, α place]	
21 -	0.411763	FV	χ-x		
22 near_close-open_mid, centralized-open_mid	0.406269	SC/ Abl	ˈɪ-ˈɛ, ˈʏ-ˈɛ, ˈʊ-ˈɔ, ˈʏ-ˈœ, ˈɪ-ˈœ, ɪ-ɛ		*e > *i / _NC, C_0i,j,u (OHG)
23 -	0.401573	PR/ SC	χ-ç	ç > x / V[-front]_	x > ç; γ > ʝ / {-V[-front]}_ (NHG)
24 approximant-fricative	0.398349	-	z-l, v-l, z-j, ʁ-j, w-v		
25 vibrant-bilabial	0.385214	-	ʀ-b, ʀ-m		
26 close_mid-diphthong	0.378769	MA	ˈeː-ˈaɪ̯, ˈoː-ˈaʊ̯, eː-aɪ̯, oː-aɪ̯		
27 open_mid-mid_close	0.368380	MA	ˈɛː-ˈeː, ɔ-o, ɛ-e, ˈɛ-ˈe		
28 approximant-plosive	0.367842	-	l-d, g-l, j-b, g-j, j-d		
29 palatal-velar	0.365951	FV	ç-x		

30 back-central	0.364320	-	ə-e, ə-a, ɪ-ɐ, ɛ-ɐ, ɐ̯-iː, ə-i, ˈɛ-ɐ̯, ɛ-ə, ˈiː-ɐ̯, ˈɪ-ɐ̯, ˈaː-ɐ̯, …		
31 affricate-fricative	0.361213	TS	t͡s-s, t͡s-f, ʃ-t͡s, ç-t͡s, t͡s-h, p͡f-f, ʃ-p͡f		
32 velar-alveolar	0.356390	→20	t-k, ŋ-n̩, ŋ-n, g-n̩, g-d, ŋ-n̩, g-n, ŋ̍-n, ŋ-d		
33 postalveolar-labiodental	0.353930	-	ʃ-f		
34 near_close-open centralized-open	0.353501	MA	ˈɪ-ˈa, ˈʏ-ˈa, ɪ-a		
35 nasal-fricative	0.353276	-	z-n̩, v-m, z-n, ʁ-m, v-n̩, v-n, ŋ-v, ʁ-ŋ		
36 voiced-semivowel	0.351379	-	ɐ̯-b, ɐ̯-l, ɐ̯-n, ɐ̯-z, ɐ̯-		

			v, e̯-d, e̯-m, l-i̯, ʀ-ɪ̯, e̯-ŋ, ʀ-i̯		
37 -	0.349631	MA	ˈiː-ˈaː		
38 nasal-plosive	0.345030	-	m-b, n-d, g-m		
39 velar-bilabial	0.337609	-	g-b, p-k		
40 glottal-labio-dental	0.330935	-	h-f		
41 semivowel-fric-ative, near_open-fricative, central-fricative	0.327919	-	e̯-f, e̯-s, e̯-h, e̯-x		
42 -	0.320832	MA	ˈiː-ˈeː		
43 postalveolar-glottal	0.320043	-	ʃ-h		
44 rounded-un-rounded	0.308460	MA	ˈʏ-ˈɪ, ˈɛ-ˈœ, ˈyː-ˈiː, ˈyː-ˈeː, ˈøː-ˈiː, yː-aː		
45 alveolar-labio-dental	0.308299	-	z-v, s-f		
46 vowel-semi-vowel	0.305100	-	ˈi̯-ˈiː, i̯-i, ˈiː-i̯, e̯-ɐ, ə-e̯, ˈɔ-e̯, e̯-uː, i̯-e		
47 glottal-alveolar	0.295420	-	ʔ-t, s-h		

48 -	0.291002	MA	ˈeː-ˈaː		
49 uvular-labio-dental	0.267084	-	ʀ-v, ʁ-v		
50 -	0.264238	MA	ˈɛː-ˈiː		
51 postalveolar-alveolar	0.240430	-	ʃ-s		
52 -	0.210080	-	ç-s		
53 semivowel-plosive, semivowel-voiceless	0.205666	-	ɐ̯-k, t-i̯, ɐ̯-t, p-i̯		
54 vibrant / uvular -approximant, uvular-palatal, vibrant-palatal	0.143678	-	ʀ-j		
55 -	0.142080	MA	eː-aː		
56	0.139879	MA	ɛ-i		
57 vibrant-plosive, uvular-plosive, uvular-velar, vibrant-velar	0.135751	-	ʀ-g		

App C.12 Phonological Rules (Derivational, Transformed Rules)

correctly identified	6	gold standard	9
free variants and doublets	5		
wrongly identified	46		
TOTAL	57		

Evaluation for the phonological rules using different threshold values. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	**Correct Sound Pair**	**Number of Identified Pairs**	**Match Precision**	**Match Recall**	**Match F-Score**
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-
0.8	0	0	-	-	-
0.75	0	0	-	-	-
0.7	0	0	-	-	-
0.65	0	0	-	-	-
0.6	1	1	1	0.111111	0.2
0.55	0	1	0.5	0.111111	0.181818
0.5	*2*	*2*	*0.75*	*0.333333*	*0.461538*
0.45	1	8	0.333333	0.444444	0.380952
0.4	2	8	0.3	0.666667	0.413793
0.35	0	11	0.193548	0.666667	0.3
0.3	0	10	0.146342	0.666667	0.24
0.25	0	4	0.133333	0.666667	0.222222
0.2	0	3	0.125	0.666667	0.210526
0.15	0	0	0.125	0.666667	0.210526
0.1	0	4	0.115385	0.666667	0.196721
0.0	0	0	0.115385	0.666667	0.196721
TOTAL	6	52			

App C.13 Old High German Sound Changes (Derivational, Transformed Rules)

Since Old High German			
correctly identified	9	gold standard	104
free variants and doublets	6	(including 1 PR)	
wrongly identified	42		
TOTAL	57		
Only Old High German			
correctly identified	2	gold standard	54
free variants and doublets	6	(including 1 PR)	
wrongly identified	49		
TOTAL	57		

Evaluation for the sound changes since (Pre-)Old High German times using different threshold values. All sound changes listed in App. A.1 are considered. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	Correct Sound Pair	Number of Identified Pairs	Match Precision	Match Recall	Match F-Score
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-
0.8	0	0	-	-	-
0.75	0	0	-	-	-
0.7	0	0	-	-	-
0.65	0	0	-	-	-
0.6	1	1	1	0.023810	0.046512
0.55	0	1	0.5	0.023810	0.045455

0.5	2	2	0.75	0.071429	0.130435
0.45	2	8	0.416667	0.119048	0.185185
0.4	*4*	*8*	*0.45*	*0.214286*	*0.290323*
0.35	0	11	0.290323	0.214286	0.246575
0.3	0	10	0.219512	0.214286	0.216867
0.25	0	4	0.2	0.214286	0.206897
0.2	0	3	0.1875	0.214286	0.2
0.15	0	0	0.1875	0.214286	0.2
0.1	0	4	0.173077	0.214286	0.191489
0.0	0	0	0.173077	0.214286	0.191489
TOTAL	9	52			

App C.14 Middle High German Sound Changes (Derivational, Transformed Rules)

Since Middle High German			
correctly identified	7	gold standard	50
free variants and doublets	6		
wrongly identified	44		
TOTAL	57		
Only Middle High German			
correctly identified	2	gold standard	12
free variants and doublets	6		
wrongly identified	49		
TOTAL	57		

Evaluation for the sound changes since (Pre-)Middle High German times using different threshold values. All Pre-Middle and Pre-New High German sound changes listed in App. A.1 are considered. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	**Correct Sound Pair**	**Number of Identified Pairs**	**Match Precision**	**Match Recall**	**Match F-Score**
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-
0.8	0	0	-	-	-
0.75	0	0	-	-	-
0.7	0	0	-	-	-
0.65	0	0	-	-	-
0.6	1	1	1	0.022727	0.044444
0.55	0	1	0.5	0.022727	0.043478
0.5	2	2	0.75	0.068182	0.125
0.45	2	8	0.416667	0.113636	0.178571
0.4	2	8	0.35	0.159091	0.21875
0.35	*0*	*11*	*0.225807*	*0.159091*	*0.186667*
0.3	0	10	0.170732	0.159091	0.164706
0.25	0	4	0.155556	0.159091	0.157303
0.2	0	3	0.145833	0.159091	0.152174
0.15	0	0	0.145833	0.159091	0.152174
0.1	0	4	0.134615	0.159091	0.145833
0.0	0	0	0.134615	0.159091	0.145833
TOTAL	7	52			

App C.15 New High German Sound Changes (Derivational, Transformed Rules)

Since New High German			
correctly identified	5	gold standard	38
free variants and doublets	6		
wrongly identified	46		
TOTAL	57		

Evaluation for the sound changes since (Pre-)New High German times using different threshold values. All (Pre)-New High German sound changes listed in App. A.1 are considered. Free variants and doublets are removed from the analysis. The threshold value with the highest F-Score is marked in italics.

Threshold (Jaccard)	Correct Sound Pair	Number of Identified Pairs	Match Precision	Match Recall	Match F-Score
0.95	0	0	-	-	-
0.9	0	0	-	-	-
0.85	0	0	-	-	-
0.8	0	0	-	-	-
0.75	0	0	-	-	-
0.7	0	0	-	-	-
0.65	0	0	-	-	-
0.6	1	1	1	0.021739	0.042553
0.55	0	1	0.5	0.021739	0.041667
0.5	1	2	0.5	0.043478	0.08
0.45	1	8	0.25	0.065217	0.103448
0.4	*2*	*8*	*0.25*	*0.108696*	*0.151515*
0.35	0	11	0.161290	0.108696	0.129870
0.3	0	10	0.121951	0.108696	0.114943
0.25	0	4	0.111111	0.108696	0.109890
0.2	0	3	0.104167	0.108696	0.106383

0.15	0	0	0.104167	0.10869565	0.10638298
0.1	0	4	0.096154	0.10869565	0.10204082
0.0	0	0	0.096154	0.10869565	0.10204082
TOTAL	5	52			

App C.16 Correct Complementary Sounds (Derivational, Transformed Rules)

Evaluation of the correctly identified complementary sounds. Feature pairs whose sounds belong to a German phonological rule or sound change are marked in italics. The first table shows the result for phonological rules, the second table for sound changes. Legend: P = positives, T = true positives, R = sonorant.

Feature Pair	Phonological Rule	P	T	Preci-sion	Recall	F-Score
1 uvular-near_open	r,ʁ,ʀ > ɐ / _C_0	2	1	0.5	1	0.666667
5 mid-syllabic	/ə/ > Ø / _C[+R]C_0	3	0	0	0	-
6 voiceless-voiced	[-R] > [-voice] / _[-R]	126	6	0.047619	0.857143	0.090226
8 short-long	V > [-long] / _[-stress]	60	3	0.05	0.428571	0.089552
20 alveolar-bilabial	N[+syllabic] > [α place] / [-continuant, -R, α place]	10	1	0.1	0.5	0.166667
23 χ-ç	/ç/ > [-front] / V[-front]_	1	1	1	1	1
TOTAL		202	12	0.059406	0.521739	0.106667

Feature Pair	Sound Change	P	T	Preci-sion	Recall	F-Score
1 uvular-near_open	ʁ > ɐ / _$	2	1	0.5	1	0.666667
5 mid-syllabic	R > R̥ / ə→0_	3	0	0	0	-
6 voiceless-voiced	D > T / _#	126	6	0.047619	0.857143	0.090226
7 back-front	*umlaut*	68	4	0.058824	0.444444	0.103896
8 short-long	V[+stress, -long] > V[+stress, +long] / _$...	60	3	0.05	0.428571	0.089552
15 plosive-fricative	x > k / _s	30	1	0.033333	1	0.064516
19 uvular-alveolar	*z > *ɹ̠ > r/ _	10	0	0	0	
22 near_close-open_mid, centralized-open_mid	*e > *i / _NC, C_0i,j,u	6	1	0.166667	1	0.285714
23 χ-ç	x > ç; γ > ʝ / {-V[-front]}_	0	0	0	0	
TOTAL		305	16	0.052288	0.484848	0.094395

App C.17 Direction of the Phonological Rules (Derivational, Transformed Rules)

Evaluation of the correctly identified direction of German phonological rules. Feature pairs with correct direction are marked in italics. Legend: “C” = most relevant condition, “VC” = *vowel_central*, R = sonorant. For the different methods, see Sect. 4.1.1.4.

Feature Pair	**Phonological Rule**	**C**	**Phonetic Plausibility**	**Articulatory Closer to the Sound Environment**	**Phonetic Distribution**
1 uvular-near_open	r,ʁ,ʀ > ɐ / _C_0	VC	*vibrant > near_open*	*vibrant > near_open*	near_open > vibrant
5 mid-syllabic	/ə/ > Ø / _C[+R]C_0	g_n	-	*syllabic > mid*	*syllabic > mid*
6 voiceless-voiced	[-R] > [-voice] / _[-R]	ˈaː_n̩	-	voiceless > voiced	*voiced > voiceless*
8 short-long	V > [-long]/ _[-stress]	_g	-	-	*long > short*
20 alveolar-bilabial	N[+syllabic] > [α place] / [-continuant, -R, α place]	k	-	*bilabial > alveolar*	*bilabial > alveolar*
29 palatal-velar	/ç/ > [-front] / V[-front]_	ˈɛ	-	velar > palatal	velar > palatal
TOTAL			1	3	4
			0.166667	0.5	0.666667

App C.18 Direction of the Sound Changes (Derivational, Transformed Rules)

Evaluation of the correctly identified direction of German sound changes. Feature pairs with correct direction are marked in italics. “C” = most relevant condition, “VC” = *vowel_central*, “FC” = *front_central*, R = sonorant. For the different methods, see Sect. 4.1.1.4.

Pair	Phonological Rule	C	Phonetic Plausibility	Articulatory Closer to the Sound Environment	Phonetic Distribution
1 uvular-near_open	ʁ > ɐ / _$	VC	*vibrant > near_open*	*vibrant > near_open*	near_open > vibrant
5 mid-syllabic	R > R̥ / ə>0_	ɡ_n	-	*syllabic > mid*	*syllabic > mid*
6 voiceless-voiced	D > T / _#	ˈaː_n̩	-	voiceless > voiced	*voiced > voiceless*
7 back-front	*umlaut*	_ç	-	*back > front*	*back > front*
8 short-long	V[+stress, -long] > V[+stress, +long] /_$...	_ɡ	-	-	*long > short*
15 plosive-fricative	x > k / _s	ˈaɪ̯_	-	-	*fricative > plosive*
19 uvular-alveolar	*z > *ɹ̝ > r/ _	FC	-	*uvular > alveolar*	alveolar > uvular
22 near_close-open_mid centralized-open_mid	*e > *i / _NC, C_0i,j,u	t͡s_	-	-	*open_mid > near_close/centralized*

29 palatal-velar	x > ç; γ > ʝ / {-V[-front]}_	ˈɛ	-	*velar > palatal*	*velar > palatal*
TOTAL			1	5	7
			0.1111111	0.5555556	0.7777778

App C.19 Conditions of the Phonological rules (Derivational, Transformed Rules)

Evaluation of the identified conditions of German phonological rules. If the direction was determined incorrectly, the condition was adjusted (column "Adj. Rule"). For more details, see Sect. 7.5.1.4. Legend: P = number of identified conditions (threshold > 0.9, see Sect. 8.1.3 and 8.2.4), TP = true positives, R = sonorant.

Pair	Phonological Rule	Adj. Rule	P	TP	Precision	Doublets
1 uvular-near_open	r,ʁ,ʀ > ɐ / _C_0	_V	68	41	0.602941	0.975610
5 mid-syllabic	R → R̥ / ə→0_	_R	10	5	0.5	0.6
6 voiceless-voiced	[-R] >[-voice] / _[-R]	-	177	27	0.152542	0.925926
8 short-long	V > [-long] / _[-stress] / ...	-	48	0	0	
20 alveolar-bilabial	N[+syllabic] > [α place] / [-continuant, -R, α place]	C[alveolar] _C_0	48	1	0.020833	0
29 χ-ç	/ç/ > [-front] / V[-front]_	V[-front]_	10	1	0.1	0
TOTAL			361	75	0.207756	0.500307

App C.20 Conditions of the Sound Changes (Paradigmatic, Transformed Rules)

Evaluation of the identified conditions of German sound changes. If the direction was determined incorrectly, the condition was adjusted (column "Adj. Rule"). For more details, see Sect. 7.5.1.4. Legend: P = number of identified conditions (threshold > 0.9, see Sect. 8.1.3 and 8.2.4), TP = true positives.

Pair	Sound Change	Adj. Rule	P	TP	Precision	Doublets
1 uvular-near_open	ʁ > ɐ / _$	_V	68	41	0.602941	0.975610
5 mid-syllabic	R >R̥ / ə>0_	_R	10	5	0,5	0,6
6 voiceless-voiced	D → T / _#		177	27	0.152542	0.925926
7 back-front	*umlaut*		20	0	0	-
8 short-long	V[-long] >V[+long] / …		48	0	0	-
15 plosive-fricative	x > k / _s		66	0	0	-
19 uvular-alveolar	*z > *ɹ̝ > r / _		24	0	0	-
22 near_close / centralized-open_mid	*e > *i / _NC, C_0i,j,u		6	1	0.166667	0
29 χ-ç	x > ç; γ > ʝ / {-V[-front]}_	V[-front]_	10	1	0,1	0
TOTAL			429	75	0.174825	0.500307
only diachronic sound changes			164	1	0.006098	

App C.21 Identified Cognates for Proto-Indo-European *reu̯H

The following list shows the aggregate PMI score for Proto-Indo-European *reu̯H 'to tear open' with all potential cognates after five iterations. The scores decrease with increasing round. Pairs with root extensions are marked with (X).

1. Round	2. Round	3. Round	4. Round	5. Round
29.84864: *reu̯p- 'to break, to tear' (X)	26.750824: *reu̯p- 'to break, to tear' (X)	23.709076: *reu̯p- 'to break, to tear' (X)	20.7271: *reu̯p- 'to break, to tear' (X)	17.772379: *reu̯p- 'to break, to tear' (X)
29.29099: *h$_3$reu̯k- 'to dig (out)'	22.686783: *h$_3$reu̯k- 'to dig (out)'	16.121422: *h$_3$reu̯k- 'to dig (out)'	9.7271: *h$_3$reu̯k- 'to dig (out)'	6.772377: *h$_3$reu̯k- 'to dig (out)'
11.594786: *u̯reh$_1$ǵ- 'to break, to tear'	3.8565102: *u̯reh$_1$ǵ- 'to break, to tear'	-2.6956844: *drep- 'to cut off, to tear off'	-4.292452: *drep- 'to cut off, to tear off'	-5.880496: *drep- 'to cut off, to tear off'
0.5947857: *drep- 'to cut off, to tear off'	-1.069901: *drep- 'to cut off, to tear off'	-3.349019: *u̯reh$_1$ǵ- 'to break, to tear'	-10.45830: *u̯reh$_1$ǵ- 'to break, to tear'	-15.71898: *u̯reh$_1$ǵ- 'to break, to tear'
-3.746143: *$^?$h$_3$u̯ei̯g- 'to open (?)'	-10.34415: *terg(^{W})- 'to tear'	-14.26801: *terg(^{W})- 'to tear'	-18.21346: *terg(^{W})- 'to tear'	-20.34713: *$^?$u̯eRǵ^h- 'to tear'
-6.372929: *terg(^{W})- 'to tear'	-10.55088: *$^?$h$_3$u̯ei̯g- 'to open (?)'	-17.19507: *$^?$h$_3$u̯ei̯g- 'to open (?)'	-18.980448 *$^?$u̯eRǵ^h- 'to tear'	-22.03759: *terg(^{W})- 'to tear'
-14.74614: *$^?$u̯eRǵ^h- 'to tear'	-16.17928: *$^?$u̯eRǵ^h- 'to tear'	-17.59524: *$^?$u̯eRǵ^h- 'to tear'	-23.53549: *peḱ- 'to pluck, to tousle'	-23.818558: *peḱ- 'to pluck, to tousle'

-21.40521: *(h$_1$)rep- ‘to snap sth. up, to pluck’	-22.94883: *peḱ- ‘to pluck, to tousle’	-23.24846: *peḱ- ‘to pluck, to tousle’	-23.89165: *$^?$h$_3$u̯ei̯g- ‘to open (?)’	-27.88050: *(h$_1$)rep- ‘to snap sth. up, to pluck’
-22.623016: *peḱ- ‘to pluck, to tousle’	-23.0699: *(h$_1$)rep- ‘to snap sth. up, to pluck’	-24.69568: *(h$_1$)rep- ‘to snap sth. up, to pluck’	-26.29245: *(h$_1$)rep- ‘to snap sth. up, to pluck’	-30.39335: *$^?$h$_3$u̯ei̯g- ‘to open (?)’
-28.1230: *k(^{W})eh$_2$d- ‘to tear, to push’	-34.0699: *(h$_1$)rei̯p- ‘to tear (down)’	-35.6957: *(h$_1$)rei̯p- ‘to tear (down)’	-37.2925: *(h$_1$)rei̯p- ‘to tear (down)’	-38.5: *ǵ/g^{h}an- ‘to yawn, to gap’
-32.4052: *(h$_1$)rei̯p- ‘to tear (down)’	-34.0700: *(h$_1$)rei̯ḱ- ‘to tear, to break’	-35.6957: *(h$_1$)rei̯ḱ- ‘to tear, to break’	-37.2925: *(h$_1$)rei̯ḱ- ‘to tear, to break’	-38.8805: *(h$_1$)rei̯p- ‘to tear (down)’
-32.4052: *(h$_1$)rei̯ḱ- ‘to tear, to break’	-36.994: *$^?$k(^{W})eh$_2$d- ‘to tear, to push’	-38.5: *ǵ/g^{h}an- ‘to yawn, to gap’	-38.5: *ǵ/g^{h}an- ‘to yawn, to gap’	-38.8805: *(h$_1$)rei̯ḱ- ‘to tear, to break’
-38.5: *ǵ/g^{h}an- ‘to yawn, to gap’	-38.5: *ǵ/g^{h}an- ‘to yawn, to gap’	-45.355: *$^?$k(^{W})eh$_2$d- ‘to tear, to push’	-53.7744: *k(^{W})eh$_2$d- ‘to tear, to push’	-60.3598: *k(^{W})eh$_2$d- ‘to tear, to push’

App C.22 Identified Cognate Pairs of the Semantic Approach

The Proto-Indo-European word pairs with the highest aggregate PMI scores (> 0.0, A. PMI) after five iterations. The open gap penalty was set to 5.5. Verb pairs with root extensions are are marked with (X).

A. PMI	Word Pair
20.330185	*ˀ*du̯er*- 'to run' : **u̯er*- 'to run' **ǵʰu̯er*- 'to walk wrily' : **u̯er*- 'to run' **u̯er*- 'to run' : **u̯erǵ*- 'to act, to make' **u̯erǵ*- 'to act, to make': **u̯er*- 'to hinder, to ward' **u̯erǵʰ*- 'to tie' : **u̯er*- 'to run'
14.830185	**dreu̯*- 'to run' : **sreu̯*- 'to flow, to stream' *ˀ*du̯er*- 'to run' : **u̯erǵ*- 'to act, to make' *ˀ*du̯er*- 'to run' : **ǵʰu̯er*- 'to walk wrily' **ǵʰu̯er*- 'to walk wrily' : **u̯erǵʰ*- 'to tie' **Hu̯er*- 'to enclose, to put in' : **u̯erǵ*- 'to enclose, to lock' (X) **reu̯H*- 'to tear open' : **reu̯p*- 'to break, to rip' **reu̯s*- 'to grub' : **dreu̯*- 'to run' **su̯er*- 'to hurt' : **u̯ers*- 'to wipe away' **tu̯er*- 'to actuate, to move sth.' : **u̯erǵ*- 'to act, to make' **tu̯erH*- 'to seize' : **u̯er*- 'to run' **u̯erǵ*- 'to act, to make': **u̯erǵʰ*- 'to tie' **u̯erǵ*- 'to act, to make': **u̯erp*- 'to turn back and forth' **u̯erǵ*- 'to act, to make': **ǵʰu̯er*- 'to walk wrily' **u̯erǵʰ*- 'to tie': *ˀ*du̯er*- 'to run' **u̯erǵʰ*- 'to tie': **u̯erǵ*- 'to enclose, to lock' **u̯erǵʰ*- 'to tie': **Hwer*- 'to enclose, to insert' **u̯ergᵘ̯*- 'to throw' : **u̯erǵ*- 'to act, to make' (X) **u̯erp*- 'to turn back and forth': **u̯ert*- 'to turn around' **u̯ert*- 'to turn around' : **u̯ers*- 'to wipe away'
11.74426	(X) **u̯reyḱ*- 'to turn, to enwind' : *ˀ*u̯reyt*- 'to turn, to twist'
9.330185	**bʰreu̯s*- 'to break' : **preu̯*- 'to jump'

	*_bʰreu̯s_- ‘to break’ : *_reu̯p_- ‘to break, to rip’ *_dreu̯_- ‘to run’ : *$^{?}$_dʰreu̯b_- ‘to drop’ *_dʰreu̯bʰ_- ‘to break into pieces’ : *_preu̯_- ‘to jump’ *_dʰreu̯bʰ_- ‘to break into pieces’ : *_reu̯p_- ‘to break, to rip’ *_ǵʰu̯er_- ‘to walk wrily’ : *_tu̯erH_- ‘to seize’ *_h_2u̯erg_- ‘to turn (around)’ : *_u̯er_- ‘to hinder, to ward’ (X) *_reu̯h_1_- ‘to open’ : *_reu̯H_- ‘to tear open’ *_tu̯erH_- ‘to seize’ : *_u̯erǵ_- ‘to act, to make’ *_tu̯erH_- ‘to seize’ : *$^{?}$_du̯er_- ‘to run’ *_u̯erǵʰ_- ‘to tie’ : *_tu̯erH_- ‘to seize’ *_u̯erh_1_- ‘to say’ : *_u̯erǵ_- ‘to act, to make’ (X) *_u̯reng_- ‘to twist, to turn’ : *$^{?}$_u̯rep_- ‘to decline’
9.119392	*_u̯er_- ‘to run’ : *_mer_-‘to disappear, to die’
6.108065	*_i̯eu̯_- ‘to hold, to tie’: *_kᵘ̯i̯eu̯_- ‘to set sth. in motion’
5.926945	(X) *_u̯er_- ‘to run’ : *_u̯reg_- ‘to follow a track’
3.830185	*_bʰreu̯H_- ‘to break sth. open’ : *_reu̯h_1_- ‘to open’ *$^{?}$_bʰreu̯k_- ‘to streak sth., to touch on sth.’ : *_preu̯s_- ‘to stray, to squirt’ *_h_2u̯erg_- ‘to turn (around)’ : *_u̯ert_- ‘to turn around’ *_h_2u̯erg_- ‘to turn (around)’ : *_u̯erp_- ‘to turn back and forth’ *_h_2u̯erg_- ‘to turn (around)’ : *_u̯ers_- ‘to wipe away’ *_reu̯H_- ‘to tear sth. open’ : *_$h_3$reu̯k_- ‘to dig (out)’ *_reu̯p_- ‘to break, to rip’ : *_$h_3$reu̯k_- ‘to dig (out)’ (X) *_u̯reyḱ_- ‘to turn, to enwind’ : *$^{?}$_u̯rengʰ_- ‘to twist together, to turn’ (X) *_u̯reyḱ_- ‘to turn, to enwind’ : *_u̯remb_- ‘to turn’ (X) *$^{?}$_u̯reyt_- ‘to turn, to twist’: *$^{?}$_u̯rengʰ_- ‘to twist together, to turn’ (X) *$^{?}$_u̯reyt_- ‘to turn, to twist’: *_u̯remb_- ‘to turn’ (X) *_u̯remb_- ‘to turn’ : *$^{?}$_u̯rengʰ_- ‘to twist together, to turn’
3.693995	*_u̯edʰ_- ‘to lead’ : *_u̯eǵʰ_- ‘to hover, to drive’ *_u̯eǵʰ_- ‘to hover, to drive’: *_u̯el_- ‘to turn, to roll’ *_u̯eǵʰ_- ‘to hover, to drive’ : *_u̯er_- ‘to hinder, to ward’ *_u̯el_- ‘to see, to recognize’ : *_u̯er_- ‘to watch, to recognize’ *_u̯es_- ‘to sell, to buy’ : *_u̯el_- ‘to turn, to roll’

	**u̯es*- ‘to graze, to eat’ : **u̯er*- ‘to run’ **u̯es*- ‘to graze, to eat’ : **u̯en*- ‘to overbear, to win’
3.619392	*$^?$*du̯er*- ‘to run’: **mer*-‘to disappear, to die’ **ǵ*^{h}*u̯er*- ‘to walk wrily’ : **mer*-‘to disappear, to die’
3.539430	**b*^{h}*er*- ‘to carry, to bring’ : **d*^{h}*er*- ‘to fix’ **b*^{h}*er* - ‘to carry, to bring’ : **u̯er*- ‘to hinder, to ward’ **b*^{h}*er*- ‘to carry, to bring’ : **ser*- ‘to take, to grasp’ **b*^{h}*er*- ‘to carry, to bring’ : **ǵ*^{h}*er*- ‘to take, to get’ **d*^{h}*er*- ‘to fix’: **ser*- ‘to arrange in rows, to link’ **ǵ*^{h}*er*- ‘to take, to get’: **per*- ‘to hit’ **ǵ*^{h}*er*- ‘to take, to get’ : **ser*- ‘to take, to grasp’ **ǵ*^{h}*er*- ‘to shine: to see’ : **u̯er*- ‘to watch, to recognize’ **mer*-‘to disappear, to die’ : **k*^{u}*er*- ‘to cut off, to carve’ **per*- ‘to hit’ : **ser*- ‘to take, to grasp’ *$^?$*red*^{h}- ‘to appear, to arise’ : **reg*^{h}- ‘to straighten up’ *$^?$*red*^{h}- ‘to appear, to arise’: **ret*- ‘to run’ *$^?$*reǵ*- ‘to flow, to drop (?)’ : **ret*- ‘to run’ **ser*-‘to look after sth./so., to protect’: **u̯er*- ‘to watch, to recognize’ **u̯er*- ‘to speak’ : **u̯er*- ‘to hinder, to ward’ **u̯er*- ‘to watch, to recognize’ : **ser*- ‘to take, to grasp’ **u̯er*- ‘to watch, to recognize’ : **ǵ*^{h}*er*- ‘to take, to get’
2.680241	**u̯ei̯s*- ‘to execute, to achieve’ : **u̯es*- ‘to sell, to buy’
2.525676	**smer*- ‘to get a portion’: **ser*- ‘to take, to grasp’
0.608065	(X) **u̯ei̯g*- ‘to get in motion, to disappear’: **u̯ei̯p*- ‘to get in swinging motion’ **u̯ei̯s*- ‘to execute, to achieve’: **u̯ei̯k*- ‘to overbear, to win’ **u̯ei̯s*- ‘to execute, to achieve’: *$^?$*tu̯ei̯*- ‘to frighten’
0.453500	**rei̯ǵ/g*- ‘to tie’: *$^?$*rei̯s*- ‘to be damaged’
0.426946	**b*^{h}*eru̯*- ‘to seethe, to bubble’ : **sreu̯*- ‘to flow, to stream’ **b*^{h}*eru̯*- ‘to seethe, to bubble’: **dreu̯*- ‘to run’ *$^?$*du̯er*- ‘to run’ : **u̯reg*- ‘to follow a track’ **ǵ*^{h}*u̯er*- ‘to walk wrily’ : **u̯reg*- ‘to follow a track’
0.426945	**u̯reng*- ‘to twist, to turn’ : **u̯er*- ‘to hinder, to ward’

App C.23 Identified Morphemes of the Semantic Approach

The following list cites examples of word pairs of the three relations identified through the semantic approach (see Sect. 5.9.3).

k-Hyponymy		**d^h-Hypernymy**		**i̯-Hypernymy**	
*d^heh_1- ‘to put; to produce, to make’	*$^?d^heh_1k$- ‘to make, to produce’	*(s)ker- ‘to clip, to scratch, to cut off’	*$^?skerd^h$- ‘to cut, to prick (?)’	*(h_1)rep- ‘to snap sth. up, to pluck’	*(h_1)rei̯p- ‘to tumble (down), to tear down’
*u̯ert- ‘to turn around’	*kert- ‘to spin, to turn’	*ḱu̯es- ‘to snuffle, to sigh’	*d^hu̯es- ‘to inhale, to exhale, to breathe’	*$^?$leu̯t- ‘to see’	*$^?$i̯eu̯t- ‘to notice, to awake’
*h_2u̯ei̯- ‘to run’	*h_2u̯elk- ‘to tow, to pull’	*$^?$ḱeu̯H- ‘to throw, to push’	*d^heu̯H- ‘to move back and forth’	*u̯eǵ^h- ‘to float; to drive’	*u̯ei̯g- ‘to start moving, to disappear’
*kers- ‘to cut (off)’	*$^?h_1$erk- ‘to cut into pieces’	*k^Wer- ‘to cut (off), to nick’	*$^?d^h$u̯er- ‘to damage, to hurt’	*$^?$kreh_2- ‘to accumulate, to collect’	*k^Wrei̯h̯- ‘to exchange, to trade’
*k^Wer- ‘to cut into pieces’	*kers- ‘to cut (off)’	*seu̯h_1- ‘to rush, to keep moving’	*d^heu̯H- ‘to move back and forth’	*$h_3$reǵ- ‘to put straight’	*rei̯ǵ- ‘to stretch, to strain’
*k^Wrei̯h_2- ‘to exchange, to trade’	*$^?$kreh_2- ‘to accumulate, to collect’	*(s)kerp- ‘to cut off, to pluck’	*$^?skerd^h$- ‘to cut, to prick (?)’	*h_2u̯elk- ‘to tow, to pull’	*h_2u̯ei̯- ‘to run’
*ser-	*snerk-	*(s)kert-	*$^?skerd^h$- ‘to cut, to prick (?)’	*u̯eǵ^h- ‘to float, to drive’	*u̯ei̯p-

‘to arrange, to link’	‘to move in together’	‘to cut (into pieces)’			‘to start swinging / trembling’
*ḱu̯ei̯t- ‘to flush up brightly’	*?ku̯ei̯t- ‘to shine’	*dreu̯- ‘to run’	*dʰeu̯- ‘to run, to hurry’	*u̯er- ‘to run’	*u̯ei̯s- ‘to flow’
*ret- ‘to run’	*?skrei̯t- ‘to go in circles (?)’	*rebʰ- ‘to move (fiercely)’	*rei̯dʰ- ‘to move swingingly’	*$h_1$eu̯- ‘to see, to catch sight of sth.’	*?i̯eu̯t- ‘to notice, to awake’
*(s)per- ‘to fly’	*(s)ḱ/ker- ‘to spring, to swing’	*kers- ‘to cut (off)’	*?skerdʰ- ‘to cut, to prick (?)’	*rebʰ- ‘to move (fiercely)’	*rei̯dʰ- ‘to move swingingly’

Appendix D: Phonotactic Approach

App D.1 Result Phonotactic Approach (Match rate)

Result list of the phonotactic method (match rate) containing the 100 sound pairs with the highest Jaccard indices and the assignment to a phonological rule (PR) of gold standard App A.2 and to a sound change of gold standard App. A.1. Doublets are in brackets. Legend: TS = Target sound, C = condition, PR = phonological rules, SC = sound change, morph. = phonotagma of a prefix or suffix, LW = phonotagma occurring in loanwords, cond. = condition of the sound change, NHG =(Pre-)New High German, MHG (Pre-)= Middle High German, OHG = (Pre-)Old High German, R = sonorant, I = i, ī, j, iu; →*NUMBER* = doublet of pair *NUMBER*.

It.	featSelect	TS	C	Type	Gold Standard 1	Gold Standard 2
1	2.453941	#	_h	PR/SC	OHG x > h / _{-$,:}	
2	2.387248	ɪ	_ç	PR/SC	NHG x > ç / {-V[-front]}_	ç > x / V[-front]_
3	2.055550	#	n_	morph.		
4	1.953079	ɛ	_ç	PR/SC	(NHG x > ç / {-V[-front]}_)	(ç>x/ V[-front]_)
5	1.868488	#	_j	PR/SC	NHG ie [ɪə̯ˀ] > je / #_	
6	1.797838	n	ə_	morph.		
7	2.177280	#	ə_	-	MHG V[-stress] > ə / _	
8	1.775461	a	_x	PR/SC	(NHG x > ç / {-V[-front]}_)	(ç>x/ V[-front]_)
9	1.723707	#	_ʃ	PR/SC	NHG s̠ > ʃ / X_t	

10	1.675262	#	_b	morph.		
11	1.619849	ʊ	_x	PR/SC	(NHG x > ç / {-V[-front]}_)	(ç>x/ V[-front]_)
12	1.640446	uː	_x	PR/SC	(NHG x > ç / {-V[-front]}_)	(ç>x/ V[-front]_)
13	1.582093	#	_f	morph.		
14	1.541129	r	_ç	PR/SC	(NHG x > ç / {-V[-front]}_)	(ç>x/ V[-front]_)
15	1.502492	a	_χ	PR/SC	(NHG x > ç / {-V[-front]}_)	(ç>x/ V[-front]_)
16	1.480486	#	ç_	-		
17	1.466136	a	ʔ_	PR/SC		Ø > [ʔ] / $(\_V\ldots)_{\omega}$
18	1.439949	#	_v	-		
19	1.402624	n	_#	morph.		
20	1.400766	t	ç_	-		
21	1.362798	ʊ	_χ	PR/SC	(NHG x > ç / {-V[-front]}_)	(ç>x/ V[-front]_)
22	1.800782	aː	_χ	PR/SC	(NHG x > ç / {-V[-front]}_)	(ç>x/ V[-front]_)
23	1.678592	ɔ	_χ	PR/SC	(NHG x > ç / {-V[-front]}_)	(ç>x/ V[-front]_)
24	2.267356	uː	_χ	PR/SC	(NHG x > ç / {-V[-front]}_)	(ç>x/ V[-front]_)
25	3.235912	oː	_χ	PR/SC	(NHG x > ç / {-V[-front]}_)	(ç>x/ V[-front]_)
26	1.335646	ɪ	_ʃ	morph.		
27	1.317747	#	ŋ_	-		g > ŋ / ŋ_
28	1.573181	k	ŋ_	PR/SC	-	n>ŋ/[-continuant, -R, +velar]_
29	1.398669	ə	ŋ_	-		
30	1.315598	aː	_x	PR/SC	(NHG x > ç / {-V[-front]}_)	(ç>x/ V[-front]_)

31	1.553415	oː	_x	PR/SC	(NHG x > ç / {-V[-front]}_)	(ç>x/ V[-front]_)
	1.553415	p	_x	-		
32	1.307501	r	iː_	-		
33	1.295111	#	ɐ_	PR/SC	NHG ʁ > ɐ / _$	r / ʀ / ʁ > ɐ / .C_0_C_0.
34	1.276123	ɛ	ʔ_	PR/SC		(Ø>[ʔ]/ (_V…)$_{\omega}$)
35	1.262650	iː	ʒ_	LW		
36	1.710815	ə	ʒ_	LW		
37	1.831008	ɛ	ʒ_	LW		
38	1.251985	#	_z	PR/SC	NHG z̥̊ > z / X_V and #_V	-
39	1.250495	n	ʧ_	LW		
40	1.542359	#	ʧ_	LW		
41	1.261031	ɛ	ʧ_	LW		
42	1.229064	n	oː_	-		
43	1.227930	a	h_	-		
44	1.217054	ʊ	_ŋ	-		
45	1.332894	a	_ŋ	-		
46	1.749292	ɪ	_ŋ	-		
47	2.459275	ɛ	_ŋ	-		
48	1.200887	t	x_	-		
49	1.194821	n	ç_	-		
50	1.144038	#	_g	morph.		
51	1.116347	#	_d	-		
52	1.113179	ɔ	ʔ_	PR/SC		(Ø>[ʔ]/ (_V…)$_{\omega}$)
53	1.109806	#	_p	-		
54	1.108463	n	_d	PR/SC	OHG t > d̥ / n_	
55	1.106264	b	yː_	morph.		

56	1.101065	#	_k	-		
57	1.099575	ɔ	_ŋ	-		
58	1.422100	g	_ŋ	→ 28		(n>ŋ/[-conti-nuant, -R, +ve-lar]_)
59	2.093490	k	_ŋ	→ 28		(n>ŋ/[-conti-nuant, -R, +ve-lar]_)
60	2.452494	ʏ	_ŋ	-		
61	1.849093	χ	_ŋ	→ 28		(n>ŋ/[-conti-nuant, -R, +ve-lar]_)
62	1.085941	#	iː_	-		
63	1.064592	t	χ_	-		
64	1.063988	ʏ	_ç	PR/SC	(NHG x > ç / {-V[-front]}_)	(ç > x / V[-front]_)
65	1.046401	aː	ʔ_	PR/SC		(Ø>[ʔ]/ (_V…)$_{\omega}$)
66	1.034905	a	_ʊ	-		
67	1.031884	ɪ	ʔ_	PR/SC		(Ø>[ʔ]/ (_V…)$_{\omega}$)
68	1.085452	ʊ	ʔ_	PR/SC		(Ø>[ʔ]/ (_V…)$_{\omega}$)
69	1.206264	iː	ʔ_	PR/SC		(Ø>[ʔ]/ (_V…)$_{\omega}$)
70	1.019413	eː	ʔ_	PR/SC		(Ø>[ʔ]/ (_V…)$_{\omega}$)
71	1.077265	øː	ʔ_	PR/SC		(Ø>[ʔ]/ (_V…)$_{\omega}$)
72	1.553084	yː	ʔ_	PR/SC		(Ø>[ʔ]/ (_V…)$_{\omega}$)
73	1.828288	uː	ʔ_	PR/SC		(Ø>[ʔ]/ (_V…)$_{\omega}$)
74	1.849093	oː	ʔ_	PR/SC		(Ø>[ʔ]/ (_V…)$_{\omega}$)
75	1.018205	t	s_	-		
76	1.138818	#	s_	PR/SC	-	z > s / _[-R]$_0$.
77	1.009492	#	x_	-		
78	1.184069	n	x_	-		
80	1.063206	#	_ʧ	LW		

81	0.993621	ɡ	ŋ_	-		
82	1.077790	l	ŋ_	-		
83	1.332718	s	ŋ_	-		
84	1.445515	ɐ	ŋ_	-		
85	0.988281	a	_p	-		
86	0.970669	#	_ʤ	LW		
87	1.321366	ɪ	_ʤ	LW		
88	1.032153	a	_ʤ	LW		
89	1.455617	ɔ	_ʤ	LW		
90	1.214618	aː	_ʤ	LW		
	1.214618	ʏ	_ʤ	LW		
91	0.964094	ɪ	ʤ_	LW		
92	0.983079	aː	ʤ_	LW		
93	0.963587	#	χ_	-		
94	1.001328	n	χ_	-		
95	1.210821	ə	χ_	-		
96	0.959243	#	t_	-		
97	0.951862	#	_ʒ	LW		
98	1.071717	aː	_ʒ	LW		
99	0.937944	ɪ	_oː	LW		
100	0.933982	#	_m	-		

App D.2 Phonological Rules (Phonotactic Method)

correctly identified	5	gold standard	26
free variants and doublets	28		
wrongly identified	69		
TOTAL	102		

Evaluation for the phonological rules after 100 iterations (in steps of 5 iterations). Doublets are removed from the analysis. The iteration with the highest F-Score is marked in in italics.

Threshold (Jaccard)	Correct Sound Pair	Number of Identified Pairs	Match Precision	Match Recall	Match F-Score
1-5	1	4	0.25	0.038462	0.066667
10	0	4	0.125	0.038462	0.058824
15	0	1	0.111111	0.038462	0.057143
20	1	4	0.153846	0.076923	0.102564
25	0	0	0.153846	0.076923	0.102564
30	1	4	0.176471	0.115385	0.139535
35	*1*	*4*	*0.190476*	*0.153846*	*0.170213*
40	0	5	0.153846	0.153846	0.153846
45	0	5	0.129032	0.153846	0.140351
50	0	5	0.111111	0.153846	0.129032
55	1	4	0.125	0.192308	0.151515
60	0	3	0.116279	0.192308	0.144928
65	0	2	0.111111	0.192308	0.140845
70	0	4	0.094340	0.192308	0.126582
75	0	4	0.094340	0.192308	0.126582
80	0	5	0.086207	0.192308	0.119048
85	0	5	0.079365	0.192308	0.112360
90	0	5	0.067568	0.192308	0.1
95	0	6	0.063291	0.192308	0.095238
100	0	5	0.031646	0.192308	0.054348
TOTAL	5	79			

App D.3 Sound Changes (Phonotactic Approach)

Evaluation for the sound changes after 100 iterations (in steps of 5 iterations). All sound changes listed in App. A.1 are considered. Doublets are removed from the analysis. The iteration with the highest F-Score is marked in italics.

Threshold (Jaccard)	Correct Sound Pair	Number of Identified Pairs	Match Precision	Match Recall	Match F-Score
1-5	3	4	0.75	0,013889	0,027273
10	1	4	0.5	0,018518	0,035714
15	1	1	0.555556	0,023148	0,044444
20	0	5	0.357143	0,023148	0,043478
25	0	0	0.357143	0,023148	0,043478
30	0	4	0.277778	0,023148	0,042735
35	1	5	0.260870	0,027778	0,050209
40	1	5	0.25	0,032407	0,057377
45	0	5	0.212121	0,032407	0,056225
50	0	5	0.184211	0,032407	0,055118
55	*1*	*5*	*0.186047*	*0,037037*	*0,061776*
60	0	5	0.166667	0,037037	0,060606
65	0	4	0.153846	0,037037	0,059701
70	0	5	0.140351	0,037037	0,058608
75	0	5	0.129032	0,037037	0,057554
80	0	5	0.119403	0,037037	0,056537
85	0	5	0.111111	0,037037	0,055556
90	0	5	0.103896	0,037037	0,054608
95	0	6	0.096386	0,037037	0,053512
100	0	5	0.090909	0,037037	0,052632
TOTAL	8	88			

App D.4 Result of the Determination of Complementary sounds (Phonotactic)

The nine correctly determined and not transposed German sound changes and phonological rules (PR/SC) after 500 iterations and the determined complementary sounds using seven different approaches (see Sect. 10.2.4): purely phonotactic (phonot.), Chafe's similarity principle (Chafe), Hoenigswald's similarity principle (Hoenig.), Dolgopolsky's classes (Dolgopol.), co-occurrent features in the sound and conditions (Condit.), subtypological principle (Subtyp.), and empirical-probabilistic principle (Empir.). The correctly determined complementary sounds are marked in italics. *R* stands for any sonorant. *It.* indicates the iteration round.

It.	PR/SC	Approach
161	r > ɐ / .C_0_C_0.	Phonot.: ʧ > ɐ / ɛː_ Chafe: h > ɐ / ɛː_ Hoenig.: ʦ > ɐ / ɛː_ Dolgopol.: d > ɐ / ɛː_ Condit.: d > ɐ / ɛː_ Subtyp.: - > ɐ / ɛː_ Empir.: d > ɐ / ɛː_
176	d > t / _#	Phonot.: l > t / _# Chafe: ʤ > t / _# Hoenig.: *d > t* / _# Dolgopol.: l > t / _# Condit.: ʦ > t / _# Subtyp.: ç > t / _# Empir.: l > t / _#
180	s̠ > ʃ / #_C	Phonot.: v > ʃ / _p Chafe: m > ʃ / _p Hoenig.: f > ʃ / _p Dolgopol.: p > ʃ / _p Condit.: f > ʃ / _p Subtyp.: h > ʃ / _p Empir.: b > ʃ / _p
257	x > k / _s	Phonot.: ɛ > k / _s Chafe: ʤ > k / _s

		Hoenig.: d > k / _s Dolgopol.: ŋ > k / _s Condit.: d > k / _s Subtyp.: ɛ > k / _s Empir.: ŋ > k / _s
273	x > ç / {-V[-front]}_	Phonot.: ʦ > ç / œ_ Chafe: r > ç / œ_ Hoenig.: z > ç / œ_ Dolgopol.: ʒ > ç / œ_ Condit.: h > ç / œ_ Subtyp.: j > ç / œ_ Empir.: ʃ > ç / œ_
335	n > ŋ / [-R,+velar]_	Phonot.: ʏ > ŋ / _k Chafe: ʏ > ŋ / _k Hoenig.: a > ŋ / _k Dolgopol.: 0 > ŋ / _k Condit.: iː > ŋ / _k Subtyp.: t > ŋ / _k Empir.: œ > ŋ / _k
381	ie > je / #_	Phonot.: n > j / _ɛː Chafe: ʒ > j / _ɛː Hoenig.: z > j / _ɛː Dolgopol.: ŋ > j / _ɛː Condit.: z > j / _ɛː Subtyp.: # > *j* / _*ɛː* Empir.: ŋ > j / _ɛː
415	0 > ʔ / $_V	Phonot.: ç > ʔ / _yː Chafe: ʦ > ʔ / _yː Hoenig.: v > ʔ / _yː Dolgopol.: p,v > ʔ / _yː Condit.: v > ʔ / _yː Subtyp.: # > ʔ / _yː Empir.: p > ʔ / _yː
446	z > s / _[-R]$_0$.	Phonot.: ç > s / _#

		Chafe: ʦ > s / _# Hoenig.: g > s / _# Dolgopol.: ʦ > s / _# Condit.: g > s / _# Subtyp.: j > s / _# Empir.: g > s / _#

App D.5 Weighted Term Frequency for Environments

The 15 most relevant conditions of [ç] in a German corpus of 1,000 sentences with three different local weight measures: absolute term frequency, tf_{ij} and tf_{2ij}. The correct environments of [ç] (i.e., after front vowels and consonants) are highlighted in italics.

Absolute *tf*		***tf**ij*		***tf**2ij*	
after_ɪ	11033	*after_ɪ*	0.367656	*after_ɪ*	0.206882
before_#	7810	before_#	0.260255	*after_œ*	0.205219
before_t	5009	before_t	0.166917	*after_ʏ*	0.171118
after_ɛ	1502	*after_ɛ*	0.050052	*after_ɛ*	0.059370
after_ʏ	1038	*after_ʏ*	0.034590	before_t	0.057316
before_ə	1008	before_ə	0.033590	*after_eː*	0.046526
after_ʁ	575	*after_ʁ*	0.019161	before_#	0.031732
after_l	438	*after_l*	0.014596	*after_ʁ*	0.016029
before_s	365	before_s	0.012163	before_ə	0.015081
before_ɐ	267	before_ɐ	0.008897	*after_l*	0.014399
after_œ	173	*after_œ*	0.005765	before_k	0.009020
before_k	130	before_k	0.004332	before_s	0.007537
after_eː	73	*after_eː*	0.002433	before_ɐ	0.006690
after_iː	72	*after_iː*	0.002399	before_oː	0.004107
after_#	63	after_#	0.002099	before_v	0.003100

App D.6 Inverse Document Frequency for Environments

The 15 most relevant conditions of [ç] in a German corpus of 1,000 sentences with three different global weight measures: inverse document frequency (*idf*), *idf-h*$_p$, and *idfm*$_p$. The correct environments of [ç] (i.e., after front vowels and consonants) are highlighted in italics.

idf		***idf-h**$_p$*		***idfm**$_p$*	
before_ʒ	3.663562	before_ʒ	3.663562	before_ʒ	2.970414
after_ʒ	2.277267	*after_œ*	3.663562	before_d	2.564949
before_ŋ	1.871802	*after_ɪ*	3.663562	before_j	2.564949
before_x	1.717651	*after_ʏ*	3.663562	*after_ɐ*	2.277267
after_œ	1.265666	*after_ɛ*	2.277267	*after_ç*	2.277267
after_ŋ	1.178655	*after_ʒ*	2.277267	before_h	2.277267
after_j	1.098612	before_t	2.277267	*after_x*	2.277267
before_ø	0.955511	*after_ɛː*	1.871802	before_z	2.054124
before_ç	0.955511	before_ŋ	1.871802	before_ʃ	2.054124
after_h	0.955511	before_x	1.717651	before_ʊ	2.054124
before_yː	0.955511	before_#	1.717651	before_f	2.054124
after_ɔ	0.890973	before_v	1.466337	*after_ʒ*	2.054124
after_yː	0.830348	before_ə	1.265666	*after_t*	2.054124
after_ɛ	0.830348	*after_øː*	1.265666	before_x	1.871802
after_z	0.830348	*after_ŋ*	1.178655	after_uː	1.871802

App D.7 Comparison of Bigrams and Trigrams

Comparison of the highest ranked phonotagms with bigrams (first and second column) and trigrams (third and fourth column), performed with the German Wiktionary dataset. The procedure was performed iteratively with a threshold of 0.5. Phonotagms of phonological rules are highlighted in italics.

Bigram	***featSelect***	**Trigram**	***featSelect***
# / _h	2.453941	n / ə_#	3.643683
ɪ / _ç	2.387248	ɪ / ʦ_oː	3.614718
# / n_	2.055550	a / v_ɪ	3.587181
ɛ / _ç	1.953079	a / h_ʊ	3.537395
# / _j	1.868488	*a / #_x*	3.503544
n / ə_	1.797838	ɪ / ɔ_ʧ	3.492731
# / ə_	2.177280	*ʊ / a_χ*	3.492731
a / _x	1.775461	b / yː_ɐ	3.486574
# / _ʃ	1.723707	ɪ / z_oː	3.467049
# / _b	1.675262	ɐ / ɛ_h	3.423921
ʊ / _x	1.619849	*k / ŋ_ə*	3.413710
uː / _x	1.640446	ɪ / n_ʃ	3.408819
# / _f	1.582093	ə / g_h	3.390003
r / _ç	1.541129	ɛ / ʦ_ɐ	3.371440
a / _χ	1.502492	ɪ / z_ʃ	3.344922
# / ç_	1.480486	r / ʊ_ʏ	3.287276
*a / ʔ_*	1.466136	a / z_ʊ	3.286880
# / _v	1.439949	*r / ʊ_ç*	3.269490
n / _#	1.402624	*a / f_χ*	3.262121
t / ç_	1.400766	r / iː_ə	3.252980

App D.8 Comparison of Iterative and Non-Iterative Procedure

Comparison of the highest ranked phonotagms with (left) and without (right) iteration. The procedure was performed with the data of App. D.7. Phonotagms of phonological rules are highlighted in italics.

Iterative	*featSelect*	Non-Iterative	*featSelect*
# / _h	2.453941	# / _h	2.453941
ɪ / _ç	2.387248	*ɪ / _ç*	2.387259
# / n_	2.055550	# / n_	2.055942
ɛ / _ç	1.953079	# / _j	1.868650
# / _j	1.868488	n / ə_	1.798597
n / ə_	1.797838	*a / _x*	1.775458
# / ə_	2.177280	# / _ʃ	1.725318
a / _x	1.775461	# / _b	1.675944
# / _ʃ	1.723707	# / _f	1.581819
# / _b	1.675262	*a / _χ*	1.498446
ʊ / _x	1.619849	# / ç_	1.481376
uː / _x	1.640446	*a / ʔ_*	1.467418
# / _f	1.582093	# / _v	1.439931
r / _ç	1.541129	n / _#	1.404497
a / _χ	1.502492	# / ŋ_	1.318873
# / ç_	1.480486	r / iː_	1.308431
*a / ʔ_*	1.466136	*# / ɐ_*	1.295304
# / _v	1.439949	iː / ʒ_	1.262069
n / _#	1.402624	# / _z	1.253162
t / ç_	1.400766	n / ʧ_	1.251757

App D.9 Comparison of Different Source Data for the Phonotactic Approach

The most relevant phonotagms after 20 iterations with untranscribed German data and a threshold set to 0.5. Five different lists and corpora were used as source data: a corpus, a lemmatized corpus, a wordform list, a lemma list (each extracted from the TIGER corpus) and a morpheme list (extracted from the CELEX data).

Corpus	Lemmatized Corpus	Wordform List	Lemma List	Morpheme List
u / q_	u / q_	u / q_	u / q_	u / q_
h / c_	h / c_	h / c_	h / c_	h / c_
# / à_	# / _j	e / _x	e / _x	k / c_
# / _j	n / _ç	h / _â	h / _â	# / _j
# / _v	o / ç_	i / q_	i / q_	ü / q_
n / _ç	# / _v	# / q_	# / q_	# / _q
e / _x	i / q_	r / è_	h / _ô	e / _x
i / q_	# / q_	h / _ô	r / è_	c / _h
# / q_	e / _x	e / v_	# / à_	# / _w
o / ã_, s / _ã, h / _â	o / ã_, s / _ã, h / _â	c / _h	e / v_	# / q_
o / ç_	# / à_	# / à_	c / _h	s / _c
r / è_	r / è_	# / _j	# / _j	e / c_
# / _à	e / d_	r / _ê, r / _ô, o / ã_, s / _ã, t / ê_, n / î_, a / _î	r / _ô, r / _ê, o / ã_, s / _ã, t / ê_, n / î_, a / _î	# / _k
# / _w	a / ç_	k / c_	n / _ç	e / g_
c / _h	# / _w	# / _v	d / _è	# / _b
a / ç_	c / _h	e / g_	k / c_	# / _p
u / ç_	e / _r	o / v_	# / _v	# / _v
# / _d	# / é_	i / v_	e / _r	a / _x
k / c_	e / b_	a / v_	o / v_	o / _x
e / b_	k / c_	n / _ç	i / v_	i / _x

App D.10 Sample of the Calculation of Articulatory Similarity of Sounds

Illustration of the calculation of articulatory similarity of sounds using the example of [d]. The same place of articulation receives "2 points"; each distance means a deduction of 0.5 points. The table below shows the calculation for vowels exemplified by [e].

	front	central	back
close	1	0.5	0
near_close	1.5	1.0	0.5
mid_close	2	1.5	1
mid	1.5	1.0	0.5
open_mid	1	0.5	0
near_open	0.5	0	0
open	0	0	0

- Bilabial: 0.5
- Labiodental: 1
- Dental: 1.5
- Alveolar: 2
- Post-alveolar: 1.5
- Retroflex: 1.0
- Alveolar-palatal: 0.5
- Palatal: 0
- Velar: 0
- Uvular: 0
- Pharyngal: 0
- Glottal: 0

Appendix E: Distinctive and Gap Approach

App E.1 Clustering with German data (all sounds)

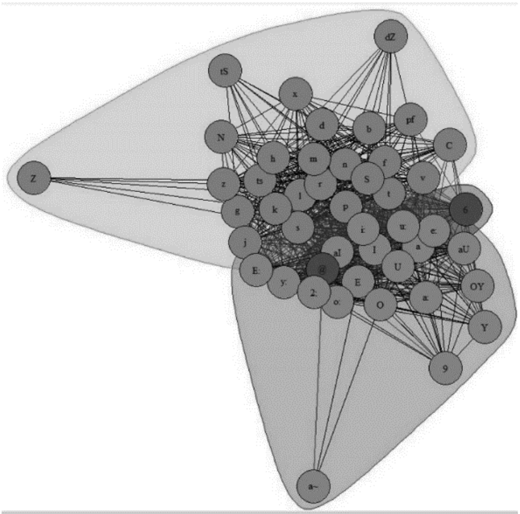

Figure App E.1: The result of clustering with German data: A network of sounds with undirected graphs weighted by the Jaccard index. A high Jaccard index between two sounds leads to spatial proximity in the network. The clusters represent consonants, vowels, and centralized vowels of unstressed syllables.

App E.2 Clustering and Final Devoicing (consonants)

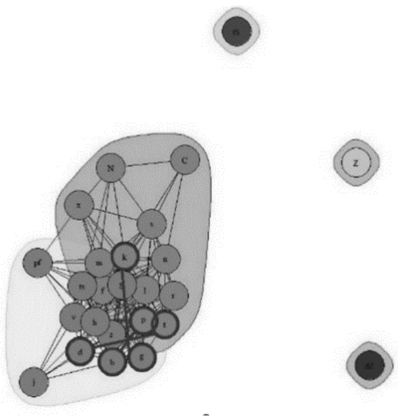

Figure App E.2: In this figure, the positions of plosives inside their clusters are marked. The voiced and voiceless plosives are assigned to different clusters through final devoicing.

App E.3 Iterative Algorithm of the Distinctive Approach

The top-ranked sound pairs sorted by JaccCos(s_1,s_2) with the German data from Fig. 11.4. The identified sound pairs are marked in italics, pairs that were removed during the iteration process are crossed out. The weights g_1 and g_2 were set to 1.0. All sounds of the same articulation type were joined to one sound class).

Rank	**Sound Pair**	**JaccCos**
1	*ʃ-ç at #_*	*1.255996*
2	*t-d at _#*	*1.243584*
~~3~~	~~ʃ-x at #_~~	~~1.243584~~
~~4~~	~~t-b at _#~~	~~1.179691~~
~~5~~	~~b-t at #_~~	~~1.151026~~
~~6~~	~~t-g at _#~~	~~1.130034~~
~~7~~	~~d-t at #_~~	~~1.023805~~
~~8~~	~~g-t at #_~~	~~1.005498~~
9	*b-p at #_*	*0.978216*
~~10~~	~~ç-ʃ at _#~~	~~0.953492~~
~~11~~	~~p-d at _#~~	~~0.942184~~
12	*z-s at #_*	*0.926230*
~~13~~	~~p-d at #C_~~	~~0.920672~~
~~14~~	~~ç-v at _#~~	~~0.919105~~
…	…	…
19	*v-x in #_*	*0.890739*
20	*k-g in _#*	*0.887901*
…	…	…

App E.4 Result of the Gap Approach

Result of the gap approach with German data. The decisive feature of the gap series is marked in italics. In the “Typo” column, all typologically probable sound changes with a value of more than 0.01 are listed. Correct sound changes are marked with TP, sound-change related gaps with wrongly identified pre-sound with WP, conditional sound change (with wrong pre-sound) with SC, sounds with single-valued features with SF, and borrowed sounds with BS.

	Pre-sound	Post-sound	SC	Typo
1	[affricate, consonant, voiced]*:[alveolar]*	[consonant, voiced, vibrant]*:[alveolar]*	*d͡z > r (SF)	0.0
2	[affricate, consonant, voiced]*:[alveolar]*	[consonant, lateral, voiced]*:[alveolar]*	*d͡z > l (SF)	0.0
3	[affricate, consonant, voiced]*:[palatal]*	[approximant, conso-nant, voiced]*:[palatal]*	*ɟ͡ʝ > j (SF)	0.0
4	[consonant, plosive voiced]*:[postalveolar]*	[affricate, consonant, voiced]*:[postalveolar]*	*ɖ > d͡ʒ (BS)	0.0
5	[rounded, back, long, vowel]*:[open-mid]*	[rounded, back, short, vowel]*:[open-mid]*	*ɔː > ɔ (WP)	0.25
6	[rounded, back, long, vowel]*:[open-mid]*	[rounded, short, vowel, front]*:[open-mid]*	*œː >œ (WP)	0.043339
7	[bilabial, consonant, voiced]*:[fricative]*	[consonant, postalveo-lar, voiced]*:[fricative]*	*β > ʒ (BS)	0.0
8	[bilabial, consonant, voiceless]*:[fricative]*	[consonant, labiodental, voiceless]*:[fricative]*	*ɸ > f (TP)	0.092199
9	[bilabial, consonant, voiceless]*:[fricative]*	[consonant, palatal, voiceless]*:[fricative]*	*Φ > ç (SC)	0.0
10	[bilabial, consonant, voiceless]*:[fricative]*	[glottal, consonant, voiceless]*:[fricative]*	*Φ > h (SC)	0.035714
11	[bilabial, consonant, voiced]*:[fricative]*	[consonant, labiodental, voiced]*:[fricative]*	*β > v (TP)	0.110429
12	[rounded, long, vowel, front]*:[open]*	[long, nasalized, un-rounded, vowel, front] *:[open]*	*ɶː > ãː (BS)	0.083562

13	[near closed, unrounded, vowel, front] :*[long]*	[rounded, closed, back, vowel]:*[long]*	*ɪː > uː (WP)	0.065359
14	[short, unrounded, vowel, front] :*[close-mid]*	[rounded, back, long, vowel] :*[close-mid]*	*e >oː (WP)	0.047014
15	[near closed, unrounded, vowel, front] :*[long]*	[rounded, closed, vowel, front]:*[long]*	*ɪː > yː (WP)	0.25
16	[rounded, long, vowel, front]:[*open*]	[short, unrounded, vowel, front]:[*open*]	*ɶː > a	0.0
17	[consonant, nasal, voiced]:[*postalveolar*]	[fricative, consonant, voiced]:[*alveolar*]	ṇ > z	0.0
18	[affricate, alveolar, consonant]:[*voiced*]	[alveolar, consonant, nasal]:[*voiced*]	d͡z > n	0.0
19	[affricate, bilabial, consonant]:[*voiced*]	[bilabial, consonant, nasal]:[*voiced*]	b͡β > m	0.0
20	[fricative, consonant, velar]:[*voiced*]	[consonant, nasal, velar]:[*voiced*]	ɣ > ŋ	0.0
21	[consonant, voiceless, velar]:[*affricate*]	[consonant, voiced, velar]:[*plosive*]	k͡x > g	0.0
22	[bilabial, consonant, stimmlos]:[*fricative*]	[alveolar, consonant, voiced]:[*plosive*]	ɸ > d	0.0
23	[consonant, plosive, voiceless]:[*postalveolar*]	[consonant, plosive, voiced]:[*bilabial*]	ṭ > b	0.0
24	[consonant, postalveolar, voiceless]:[plosive]	[bilabial, consonant, voiceless]:[affricate]	ṭ > p͡f	0.0
25	[consonant, plosive, voiceless]:[postalveolar]	[affricate, consonant, voiceless]:[postalveolar]	ṭ > t͡ʃ	0.0

Eo Ipso – Automated Internal Reconstruction

This volume aims to contribute to the study into the method of internal reconstruction of ancient language stages in two particular ways. Firstly, it seeks to give texture to the theoretical aspects of internal reconstruction, including its problems and basic requirements. The second part deals with the machine implementation of internal-reconstruction methods. Different methods that have been proposed in the literature are presented and their suitability for an adequate computational implementation is discussed. The emphasis is set on the phonological reconstruction of sound change. In total, six different methods are implemented with German and Proto-Indo-European data. The results of the automated internal reconstruction are evaluated quantitatively and qualitatively against a gold-standard.

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

TÜBINGEN LIBRARY PUBLISHING

ISBN 978-3-946552-74-1

9 783946 552741