Background on Comparative Linguistics

Johann-Mattis List (University of Passau)

1 Preliminary Considerations

What is a Language?

What counts as a languages, i.e. which tradition of speech we label as language, does not depend on pure linguistic criteria, but also on social and cultural criteria (Barbour and Stevenson 1998: 8). Accordingly, we assume that people in Shànghǎi, Běijīng, and Měixiàn all speak dialects of "Chinese", while people in Scandinavia speak languages such as "Norwegian", "Swedish", or "Danish". This does not mean that the Chinese varieties show less differences than the Scandinavian ones, as we can see from Table 1:

Běijīng Chinese	1	iou²¹ i⁵⁵	xuei³⁵	pei²¹fən⁵⁵	kən⁵⁵	tʰai⁵¹iaŋ¹¹	t͡ʂəŋ⁵⁵ tsai⁵³	naɚ⁵¹	t͡ʂəŋ⁵⁵luən⁵¹
Hakka Chinese	1	iu³³ it⁵⁵	pai³³a¹¹	pet³³fuŋ³³	tʰuŋ¹¹	ɲit¹¹tʰeu¹¹	hɔk³³	e⁵³	au⁵⁵
Shànghǎi Chinese	1	ɦi²²	tʰã⁵⁵ tsɿ²¹	poʔ³foŋ⁴⁴	taʔ⁵	tʰa³³ɦiã⁴⁴	tsəŋ³³ hɔ⁴⁴		ləʔ¹lə²³tsa⁵³

Běijīng Chinese	2	ʂei³⁵		də⁵⁵		pən³⁵ liŋ²¹	ta⁵¹	
Hakka Chinese	2	man³³	ɲin¹¹		kʷɔ⁵⁵	vɔi⁵³		
Shànghǎi Chinese	2	sa³³	ɲiŋ⁵⁵	ɦiəʔ²¹		pən³³ zɿ⁴⁴	du¹³	

Norwegian	1	nuːɾɑʋinˑɳ	ɔ	suːlɳ				kɾɑŋlət	ɔm
Swedish	1	nuːɖɑnʋɪndən	ɔ	suːlən		tyɪstadə	ən gɔŋ		ɔm
Danish	1	noʌʌnvenˀɳ	ʌ	soːlˀn	kʰʌm		eŋ̊ɑŋ	i sɖʁiðˀ	ʌmˀ

Norwegian	2	ʋem	ɑ	dem	sm̩	ʋɑː	d̩ɳ	stæɭkəstə	
Swedish	2	ʋɛm	ɑʋ	dɔm	sɔm	ʋa		staɹkast	
Danish	2	ʋɛmˀ	a	b̥m̩	d̥	ʋa	d̥ɳ	sɖæʌɡ̊əsɖə	

Table 1: "Der Nordwind und die Sonne" in verschiedenen Sprachvarietäten

---

The table shows phonetic transcriptions of the translation of the sentence "The Northwind and the sun were disputing, who was stronger" in six different linguistic varieties. Unfortunately, there is no further information on the structure of the table. How can we explain it anyway? Which conclusions can be drawn with respect to the classification of Chinese speech varieties into dialects and Scandinavian speech varieties into languages?

Language as a Diasystem

In order to allow linguists to handle the complex, heterogeneous character of languages more realistically, sociolinguistics usually invokes the model of the *diasystem* (Bussmann 1996: 312). According to this model, languages are complex aggregates of different linguistic systems, which 'coexist and influence each other' (Coseriu 1973: 40).[1] An important aspect is the existence of a so-called "roof language" (*Dachsprache*), i.e., a language variety which serves as standard for interdialectal communication (Goossens 1973: 11). The linguistic varieties (dialects, sociolects) which are connected by such a standard constitute the "variety space" (*Varietätenraum*) of a language (Oesterreicher 2001), as shown in Figure 1.

---

[1]My translation, original text: "die miteinander koexistieren und sich gegenseitig beeinflussen"
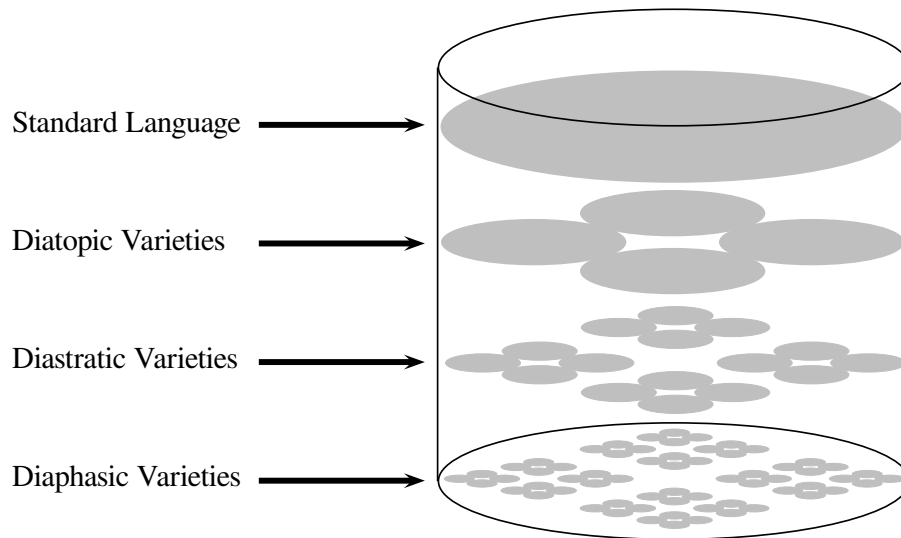
Figure 1: Language as a diasystem

How can the model of the diasystem help us to explain the different division of Chinese and Scandinavian speech varieties into dialects and languages?
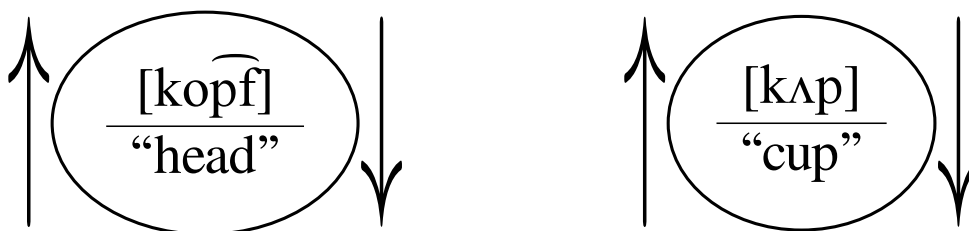
What is a Linguistic Sign?

In historical linguistics, linguistic signs are usually treated in the context of the traditional sign model by Saussure (*Cours de linguistique générale*). As Roman Jakobson notes, we distinguish two sides: the form and the content:

> The sign has two sides: the sound, or the material side on the one hand, and meaning, or the intelligible side on the other. Every word, and more generally every verbal sign, is a combination of sound and meaning, or to put it another way, a combination of signifier and signified [...]. (Jakobson 1976 [1978]: 3)

What does Jakobson mean with the words "material" and "intelligible"?

An Extended Sign Model for Comparative Linguistics

Normally, the classical sign model by Saussure is depicted as follows:

$$\frac{[\widehat{kopf}]}{\text{``head''}} \qquad\qquad \frac{[k\Lambda p]}{\text{``cup''}}$$

Important for the linguistic sign is, however, not only the *form* (signifier) and the *meaning* (signified), but also the linguistic *system* in which the sign is used. A more detailed depiction of the sign model should therefore also include the system as a constitutive aspect of the linguistic sign:
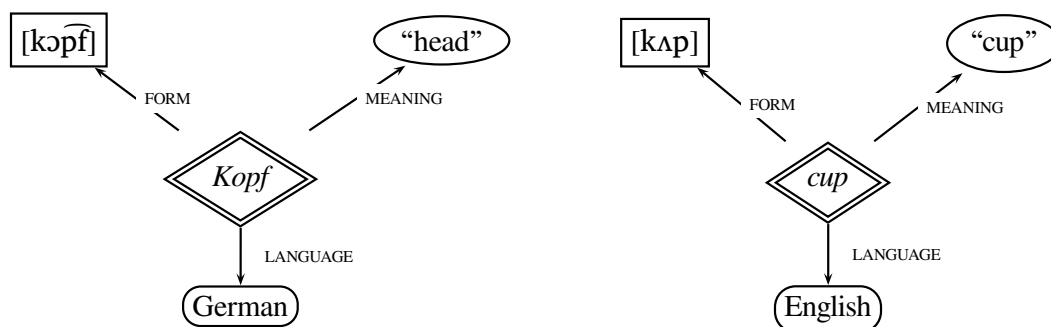
> If we look at the structure of sign form and sign meaning, we can find fundamental differences between the two. The sign form is a (phonetic) sequence, that is, a linear arrangement of distinctive sounds. These sounds are material, since they can be measured as waves in the air, or as traces of ink on a sheet of paper. Important for the sign form is furthermore its linearity, since not only the assembly of different sounds is crucial for the distinction between different sign forms, but also the order of elements. We can therefore say that the sign form is (a) **substantial**, (b) **segmentable**, and (c) **linear**. But what about the sign meaning? Fill in the corresponding terms in the right column of the table.

No.	Form	Meaning
(a)	substantial	
(b)	segmentable	
(c)	linear	

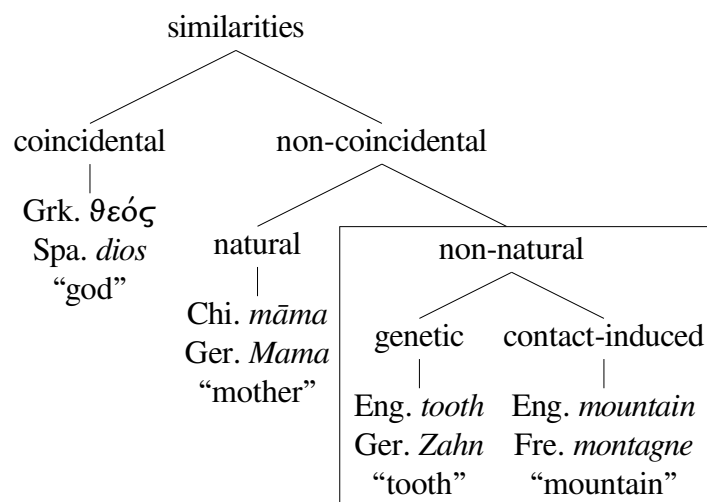### How do we Compare Languages?

In a very simple model, we can say that a language consists of a certain number of words (or linguistic signs, as we have seen before) and a certain number of syntactic rules by which these words can be combined to form phrases. In spoken languages, the words themselves are formed from a fixed number of sounds which can be combined according to a fixed number of phonotactic rules.

While this model of language as a bag of words may seem very simple, it is effectively the model that was underlying most of the quantitative comparative analyses that have been published so far. Additionally one should say, that even classical linguists who do not work in a quantitative framework tend to use this model in their analyses.

When comparing languages, we need to identify a *tertium comparationis*, that is, we need to find aspects according to which we compare languages. Similar to comparing two objects, for example, two bicycles, we will try to break down the comparison to certain *features*, such as the wheels of our bikes, or their saddle. By comparing the characteristics of these features, e.g., the size of the wheels, or their thickness, we can then start to draw certain conclusions.

As a very simple conclusion, we could try to determine if the bikes are from the same brand. But we can also ask, whether they have been built for the same purpose, or whether they are used in similar environments. These three factors do not need to coincide, and one may need to be an expert in bike construction to learn more about it, but whenever we compare objects with each other, we essentially (1) identify certain similarities based on certain *comparative concepts* (Haspelmath 2010) which serve as the basis of our comparison, and we can then (2) seek explanations for the similarities between the objects.

> When only considering similarities between words, we can see four different kinds of similarities presented in the following figure (based on List 2014). How do these similarities relate to our bicycle example, and how do they relate to comparative linguistics and its sub-disciplines?

```
                                similarities
                              /              \
                   coincidental            non-coincidental
                        |                    /            \
                   Grk. θεός            natural        non-natural
                   Spa. dios               |          /          \
                    "god"             Chi. māma    genetic    contact-induced
                                      Ger. Mama       |             |
                                       "mother"   Eng. tooth   Eng. mountain
                                                  Ger. Zahn    Fre. montagne
                                                   "tooth"      "mountain"
```

2 Historical Linguistics

Objective

One of the core objectives of investigating languages from a historical viewpoint is to find out how they *evolved* into their current shape. Similarities of interest for historical linguistics are therefore always those similarities that can be shown to be a result of common ancestry. Since language change goes peculiar pathways, it may not always be easy to find a proper *tertium comparationis* in historical linguistics. What surfaces as an article in one language may well go back to an older demonstrative and surface as a copula in another language. For this reason, the primary focus of historical linguists in identifying historical similarities between languages is not the function or the meaning of a given word or morpheme in a given language, but the sounds from which these are built. Although sounds also change their shape, it has been convincingly shown that they do so in a rather systematical manner. Therefore, when finding the patterns underlying the correspondences of sounds across different languages, it is often rather easy to determine if the languages are historically related and how closely.

> The description of objectives given above does not provide any further information on the areas where historical linguists investigate language evolution. Which ones are probably the most important areas (or aspects of language) in which historical linguists investigate how change proceeds?

Methods

The apparently most important method employed in historical linguistics is the so-called *comparative method*. The comparative method is an overarching framework that historical linguists use to study language history. The application of the framework is tedious, involving many iterative steps. Scholars start by comparing words from different languages in order to identify sets of potentially related words (*cognates*). They then set up lists of sound correspondences and use this information to revise their initial list of cognates (see Table 3). This new information is again used to revise the list of corresponding segments, and so on, until the results can no longer be refined. By applying this method to two or more languages, linguists assemble *cognate words* and *correspondence patterns*, which are then used to infer change scenarios that explain the different correspondence patterns by invoking an ancestral language from which the sounds in the descendant languages (the reflex sounds) can be derived in the most convincing fashion.

Apart from the comparative method, historical linguists have developed and are developing additional methods to handle different topics, such as, for example, semantic change (which we will discuss in Session 3), but also the topic of *phylogenetic reconstruction* enjoys some prominence, although some scholars subsume the classical, non-computational techniques under the framework of the comparative method itself.

> The table below gives an example with respect to the detection of sound correspondences between English and Ancient Greek. How can the principle be handled for more than one language?

Cognate List		Alignment					Correspondence List		
							Eng.	Grk.	Freq.
English	*foot*	f	ʊ	t			f	p	3 x
Ancient Greek	πoδ-	p	ɔ	d			f	pʰ	1 x
English	*father*	f	ɑː	θ	ə	ɹ	ɹ	r	2 x
Ancient Greek	πατέϱ-	p	a	t	ɛ	r	θ	t	1 x
English	*fear*	f	ɪə	ɹ	-		t	d	1 x
Ancient Greek	φoβέ-	pʰ	ɔ	b	e		*irregular*		
English	*fire*	f	aɪə	ɹ			*match!*		
Ancient Greek	πυϱ-	p	y	r					

Detecting regular sound correspondences in classical historical language comparison.

Models

Scholars like Jacob Grimm had a rather fuzzy understanding of the historical relatedness of languages, and many scholars kept thinking that contemporary languages could be directly "derived" from each other. This changed in the mid of the 19th century, when scholars started to take the idea that languages seem to evolve in tree-like patterns more seriously. While this idea had been around for some time before the advent of "modern" historical linguistics (List et al. 2016), it was not until scholars like August Schleicher (1821-1868) started to propagate the idea not only in words, but also in illustrations (Schleicher 1853, Schleicher 1861), that the family tree model of language history was accepted as something useful to discuss in historical linguistics.

By now, the family tree can be seen as one of the most influential models in historical linguistics. Although it has been challenged, language evolution can hardly be studied without it. The same cannot be said about models for sound change or semantic change. While these models exist, they are much more detailed and specific and rarely gain such a huge acceptance as the tree model of language diversification.

Figure 1: Schleicher's early tree from 1853, and an attempt to visualize the wave theory by Schmidt (1875).

> If you compare Schleicher's early tree drawing from 1853 with modern phylogenetic trees, they will look quite different, in terms of abstraction. What could this reflect about the thoughts of the authors?

3 Linguistic Typology

Objective

While historical linguistics deals with the development of particular languages or language families, linguistic typology focuses on those aspects of languages which surface independently of individual language histories. While historical linguistics concentrates on those similarities among languages which are due to change among particular languages, linguistic typology seeks to identify those similarities which have developed independently from a languages' descent. Following our comparison with bicycles, linguistic typology would be interested in the various types of bikes which are being produced (e.g., mountain bikes, road bikes, etc.), while historical linguistics is interested in brands.

> At times it appears that linguistic typology deals with synchrony while historical linguistics deals with diachrony. Is this reasonable?

Methods

There are multiple ways of comparing languages, and there is a large number of aspects for which languages can be compared. Given that – unlike historical linguistics – typology deals with more abstract similarities that are not due to common descent, it is more difficult to find suitable *tertia comparationis*, or *comparative concepts*, as they are called by Haspelmath (2010). In typology and in linguistics in general, there is a rather heated debate about the nature of the comparative concepts that linguists define and select in order to compare different languages with each other. A concept like *case*, for example, can be interpreted in multiple ways, and it is not always clear how case should be understood. The confusion also arises from tradition. The Latin *ablative* case, for example, is not a true ablative in the original sense of the word, denoting a case that indicates the starting point of a departure, answering the question "from where", as it is still the predominant usage of the ablative case in Sanskrit. Instead, the Latin ablative shares many properties with the Russian *instrumental* case, which itself is not a true instrumental anymore, as it is again used to express many additional functions that are not predominantly related to the instrumental use of a given object, answering the question "with what?". When starting from the semantics, on the other hand, for example from the questions which are taught in school times in order to deal with case in inflecting languages like Latin, it is clear that languages use different strategies to encode the relevant information, and some could belong to some general grammatical notion of *case*, while other strategies are also available and actively used by many of the world's languages.

But the debate goes beyond pure terminology, since typologists often do not agree with respect to the reality behind the comparative concepts they use. Some linguists say they reflect (or should reflect) some deep innate properties that might find their direct reflection in our brains, some say they are mere tools for comparison, which may be practically defined, but do not need to have a clear relation to any deeper reality, and some scholars take an intermediate position, emphasizing that some of the concepts by which linguists compare languages are useless, but that there should be some deeper value to them. Haspelmath (2018), for example, emphasizes that there is a crucial distinction between language-specific categories, such as the *ablative* in Latin, and cross-linguistic comparative concepts, but that linguists often confuse the two, since they wrongly assume that linguistic categories would have a direct manifestation similar to the idea of *natural kinds* in physics and chemistry. Bond (2019)

and other proponents of *Canonical Typology*, on the other hand, argue that cross-linguistic comparison can be carried out by relying on the notion of a *canon*, that is, a "logically motivated archetype from which attested and unattested patterns are calibrated" (Bond 2019: 83).

No matter how typologists motivate their comparative concepts in the end, it seems clear that the techniques which have been developed to compare languages typologically have greatly improved during the last decades and centuries. As a result, language comparison is nowadays much less biased towards classical European languages and Sanskrit than it was before.

> Why does semantics play such an important role in typological language comparison?

Models

While historical linguistics has a standard model of language evolution, we do not find comparable standard models of language typology in the field of linguistic typology. The reason for a lack of unified models is that it is extremely likely that there is no unique reasons for similarity across languages which are not due to contact or common descent, but rather an interaction of multiple factors. Common factors mentioned and investigated by linguists include (1) efficiency of coding (Nettle 1995), (2) climate (Everett et al. 2015), (3) population size (Bromham et al. 2015), or (4) social structure (Lupyan and Dale 2010).

> Judging from the short list of only four factors mentioned here, why is it clear that these are not necessarily competing models of linguistic typology?

4 Areal Linguistics

Objective

While languages can be similar due to common descent or due to general properties that all human languages share, there is a third non-trivial reason why languages can exhibit similarities: language contact. In contact situations, when there is a sufficient number of bilingual speakers, not only words but also structures can be easily transferred from one language to another. To identify which material can be transferred during contact, and under which circumstances and with which dynamics language contact occurs can be seen as the primary objective of *areal linguistics*.

> In the bicycle example above, it was mentioned that bikes can be similar when they are used in similar environments. Does this reflect a situation similar to language contact?

Methods

We have already seen that it is rather difficult to say exactly what the methods are which are used in linguistic typology, which is why we looked at the selection of comparanda, or comparative concepts, rather than discussing specific methodological frameworks. In areal linguistics, we have similar problems, since it is difficult to identify a unified methodological framework. Instead, scholars use different shortcuts in order to distinguish borrowed from non-borrowed traits (see the short overview in List 2019).

> Could the above-mentioned comparative method be used for lexical comparison in the realm of areal linguistics?

Models

At times, scholars contrast the model of a family tree in historical linguistics with the wave model in areal linguistics. The major idea is that innovations, that is, novel ways of speaking, can expand across dialect continua and contact areas in form of waves that may not reach all corners of a given area. What a wave cannot model that well, however, is the direction of influence, and specifically in those cases where we can find many borrowings between languages in well-known contact areas, such as South-East Asia, we find that languages do not influence each other mutually, but that often one language may exhibit more influence over another language. Here, a model of a directed network seems to be much more useful to model contact phenomena.

<div style="border:1px solid">

What is a directed network?

</div>

References

Barbour, S. and P. Stevenson (1998). *Variation im Deutschen. Soziolinguistische Perspektiven*. Berlin: de Gruyter.

Bond, O. (2019). "Canonical Typology." In: *The Oxford handbook of morphological theory*. Ed. by J. Audring and F. Masini. Oxford: Oxford University Press, 409–431.

Bromham, L., X. Hua, T. G. Fitzpatrick, and S. J. Greenhill (2015). "Rate of language evolution is affected by population size." *Proceedings of the National academy of Sciences of the United States of America* 112.7, 2097–2102.

Bussmann, H., ed. (1996). *Routledge dictionary of language and linguistics.* Trans. from the German by G. Trauth and K. Kazzazi. London and New York: Routledge.

Coseriu, E. (1973). *Sincronía, diacronia e historia. El problema del cambio lingüístico* [Synchrony, diachrony, and history. The problem of linguistic change]. Madrid: Biblioteca Románica Hispánica.

Everett, C., D. E. Blasi, and S. G. Roberts (2015). "Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots." *Proceedings of the National Academy of Sciences of the United States of America* 112.5, 1322–1327.

Goossens, J. (1973). *Niederdeutsch. Sprache und Literatur. Eine Einführung*. Neumünster: Karl Wachholtz.

Haspelmath, M. (2010). "Comparative concepts and descriptive categories." *Language* 86.3, 663–687.

— (2018). In: *Aspects of linguistic variation*. Ed. by D. V. Olmen, T. Mortelmans, and F. Brisard. Berlin and New York: De Gruyter Mouton, 83–113.

Jakobson, R. (1976 [1978]). *Six lectures on sound and meaning*. Trans. from the French by J. Mepham. With an intro. by C. Lévi-Strauss. Cambridge and London: MIT Press.

List, J.-M. (2014). *Sequence comparison in historical linguistics*. Düsseldorf: Düsseldorf University Press.

— (2019). "Automated methods for the investigation of language contact situations, with a focus on lexical borrowing." *Language and Linguistics Compass* 13.e12355, 1–16.

List, J.-M., J. S. Pathmanathan, P. Lopez, and E. Bapteste (2016). "Unity and disunity in evolutionary sciences: process-based analogies open common research avenues for biology and linguistics." *Biology Direct* 11.39, 1–17.

Lupyan, G. and R. Dale (2010). "Language structure is partly determined by social structure." *PLoS ONE* 5.1, e8559.

Nettle, D. (1995). "Segmental inventory size, word length, and communicative efficiency."

Oesterreicher, W. (2001). "Historizität, Sprachvariation, Sprachverschiedenheit, Sprachwandel." In: *Language typology and language universals. An international handbook*. Ed. by M. Haspelmath. Berlin and New York: Walter de Gruyter, 1554–1595.

Saussure, F. de. *Cours de linguistique générale*. Ed. by C. Bally. Lausanne: Payot, 1916; German translation: — . *Grundfragen der allgemeinen Sprachwissenschaft*. Trans. from the French by H. Lommel. 2nd ed. Berlin: Walter de Gruyter & Co., 1967.

Schleicher, A. (1853). "Die ersten Spaltungen des indogermanischen Urvolkes ᴛʜᴇ ꜰɪʀꜱᴛ ꜱᴘʟɪᴛꜱ ᴏꜰ ᴛʜᴇ ɪɴᴅᴏ–ᴇᴜʀᴏᴘᴇᴀɴ ᴘᴇᴏᴘʟᴇ." *Allgemeine Monatsschrift für Wissenschaft und Literatur* 3, 786–787.

— (1861). *Compendium der vergleichenden Grammatik der indogermanischen Sprache*. Vol. 1: *Kurzer Abriss einer Lautlehre der indogermanischen Ursprache*. Weimar: Böhlau.

Schmidt, J. (1875). *Zur Geschichte des indogermanischen Vocalismus. Zweite Abteilung*. Weimar: Hermann Böhlau.