

Mouths, tongues, and ears

Source concepts for 'language' across Africa

 **Alice Mitchell** | University of Cologne

Nicola Zimmermann | University of Cologne

 <https://doi.org/10.1075/clu.23.04mit>

 Available under a CC BY-NC-ND 4.0 license.

Pages 83–103 of

Anthropological Linguistics: Perspectives from Africa

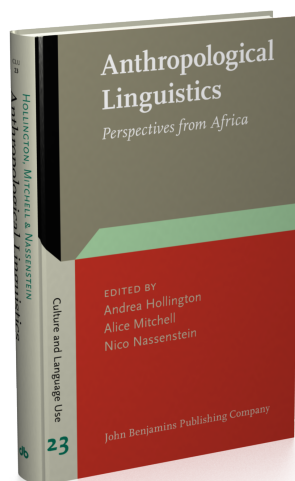
Edited by Andrea Hollington, Alice Mitchell and Nico Nassenstein

[*Culture and Language Use*, 23] 2024. xiii, 485 pp.

© John Benjamins Publishing Company

This electronic file may not be altered in any way. For any reuse of this material, beyond the permissions granted by the Open Access license, written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

For further information, please contact rights@benjamins.nl or consult our website at benjamins.com/rights



Mouths, tongues, and ears

Source concepts for ‘language’ across Africa

Alice Mitchell & Nicola Zimmermann

University of Cologne

This paper investigates words for ‘language’ and their possible source concepts in 130 African languages. Radden (2001) noted that words for ‘language’ almost always draw on more basic meanings and proposed a metonymic chain from (i) speech organs to (ii) speaking as activity to (iii) speech as “object” to (iv) language. Our sample provides additional evidence for these semantic pathways, showing that ‘tongue’ and ‘mouth’ are important source concepts for ‘language’ across Africa, as well as concepts relating to speech. After summarizing the survey results, we consider the novel extension ‘ear’ to ‘language’ in Datooga. While frequently attested metonymic processes point to widespread conceptualizations of language as embodied, we nonetheless find cultural variation in which aspects of speech behavior provide conceptual material for words for ‘language’.

Keywords: language, metonymy, semantic typology

1. Introduction

The concept ‘language’ is central to our work as linguists. For many, ‘language’ is also a troublesome concept: we recognize that notions of distinctive languages, of bounded linguistic systems, are ideologically as well as analytically motivated (Irvine & Gal, 2000; Lüpke & Storch, 2013). While linguists have been thinking carefully about how to categorize and abstract from linguistic behavior – by focusing on repertoires, registers, and codes, and experimenting with terms like translanguaging and heteroglossia (e.g., Wei, 2018), in everyday communication people continue to operate with concepts of ‘language’ that revolve around the idea of a distinctive system or medium of communication. Where do these everyday concepts come from? In this paper, we address this question from the perspective of lexical semantics, comparing words for ‘language’ across Africa and

considering the conceptual sources for this abstract concept. As we will see, we find a great deal of polysemy in this semantic domain.

The paper focuses on nominal forms that denote an abstract system of communication. In many cases, these words for ‘language’ can be modified with labels referring to communities defined along ethnic or geographic lines, e.g., *language in the English language* or *úttá in úttá Dátóoga* ‘the Datooga language’. Of course, the existence and utility of the concept ‘language’ emerges from a specific set of socio-historical conditions, often related to territorial expansion, conflict, and colonialism, in the process of which people come to distinguish ways of speaking. Racist and other prejudiced ideologies of language also play a role in deciding which varieties get labeled as ‘languages’ at all (as opposed to ‘dialects’). Goddard (2011) links the concept of distinctive languages with political processes of standardization in Europe and notes that many languages do not have an equivalent lexicalized meaning.¹ We consulted several dictionaries of African languages that had no entry for ‘language’ and we acknowledge the historically contingent nature of this concept.² Much work is still to be done to better understand cultural notions of language in Africa and the extent to which the term ‘language’, as used by academics, is even appropriate (Makoni & Mashiri, 2007). Nonetheless, contact between groups of people speaking different linguistic varieties as well as an appreciation of diverse ways of speaking is the norm in Africa, and was so long before the colonial period. Consequently, we expected that a comparative study of words for ‘language’ would prove fruitful and indeed we were able to collect over 150 lexical items from different languages that contained ‘language’ as part of their dictionary definition.

Despite linguists’ preoccupation with language, very little cross-linguistic research has explored the lexicalization of the concept ‘language’, with the important exception of Radden (2001), whose work we discuss in detail below. Goddard (2011) offers an in-depth analysis of the meaning of ‘language’ in English, with some reference to other languages. Another small-scale cross-linguistic study is Newman and Schuh (2016), which examines differences in the formation of language names in Hausa compared with other Chadic languages. They show that while Chadic languages typically extend the word ‘mouth’ to mean ‘language’, Hausa uses the word ‘tongue’ and has also innovated a derivational suffix that

1. Partly on this basis, ‘language’ is not a semantic prime in the Natural Semantic Metalanguage, though the speech act ‘say’ is, as well as the concept ‘words’ (Goddard & Wierzbicka, 2014).

2. For example, the trilingual Soninké (Mande; Mali)-English-French dictionary has no entry for ‘language’, while the entry for *langue* gives the Soninké form *neene*, which has several meanings, including the organ ‘tongue’, but not ‘language’. Online dictionary available at https://www.asawan.org/fr/lexique_sonink%C3%A9, accessed 23 November 2020.

combines with geographic or ethnic labels to denote the language of the respective group. (Variation in source concepts for ‘language’ within language families is briefly addressed in Section 4, where we consider divergent patterns of polysemy in the Nilotic family.) Though people’s words for ‘language’ have not been much studied in linguistics, various researchers have investigated linguistic expressions relating to speech and to manners of speaking, which, like words for ‘language’, frequently involve metonymic use of body part terms such as ‘tongue’, ‘mouth’, or ‘throat’. Sweetser (1992) briefly considers references to the physical articulators in English metaphors like “the words stuck in my throat” (1992, p.712) as well as discussing the importance of the conduit metaphor (i.e., passing objects back and forth) in how we talk about communication. Goosens (2009) identifies various semantic source domains for metaphorical language about linguistic interaction in English, including body parts, sounds, and violent actions. Metonymic use of the body part terms ‘mouth’, ‘lips’, ‘tongue’ and ‘teeth’ has been shown to be significant in talking about verbal behaviour in Chinese (Jing-Schmidt, 2008) as well as Bulgarian (Bagasheva, 2017). While the idiomatic expressions speakers of different languages use to talk about communicative styles are highly relevant for studying cultural conceptualizations of language, here we restrict our focus solely to lexical items meaning ‘language’ and their conceptual sources, as previously investigated by Radden (2001).

Based on the observation that few languages have a basic word for ‘language’, Radden (2001) investigates the source domains for this concept across a small, random sample of the world’s languages. His goal is to find “cross-linguistic evidence for a folk model of language” (2001, p.57). He considers both synchronic and diachronic semantic relations and identifies three main source domains for ‘language’: (i) articulation and speech organs (e.g., ‘mouth’, ‘tongue’); (ii) linguistic action (e.g., ‘speech’, ‘talk’); and (iii) basic linguistic units (e.g., ‘word’). These domains can all be linked to abstract notions of ‘language’ via metonymic processes. Radden proposes a metonymic chain from concrete speech organs to an abstract concept of ‘language’ via the act of speaking and the ‘results’ of speech, and he relates these language-specific metonymies to more general ones, as presented in Table 4.1.

This paper contributes more empirical data to testing Radden’s suggestion that “[p]ossibly the ways we speak about language are very similar cross-linguistically” (2001, p.57). Because we are reliant on dictionaries, we focus on synchronic patterns of polysemy in words for ‘language’, rather than diachronic semantic developments. As discussed in Section 3, we find substantial additional evidence for the metonymies presented in Table 4.1, but we also identify a range of other semantic relations. One unique metonymy found in the Nilotic language Datooga is EAR FOR LANGUAGE, which we analyze in some detail in Section 4.

Table 4.1 Metonymic chain from ‘speech organ’ to ‘language’, adapted from Radden (2001)

speech organ	>	speaking	>	speech	>	language
specific		SPEECH ORGAN FOR		SPEAKING FOR		SPEECH FOR
metonymies		SPEAKING		SPEECH		LANGUAGE
general		INSTRUMENT FOR		ACTION FOR		SPECIFIC FOR
metonymies		ACTION		RESULT		GENERIC

While this metonymy can be loosely accommodated into Radden’s model, his claim that the articulatory organs are basic to all human concepts of ‘language’ must be modified to include organs of speech perception as well as production. Similarly, ‘speaking’ is not the only linguistic activity that provides conceptual material for ‘language’: ‘hearing’ can play a role here too.

As with Radden’s paper, the motivation behind the present survey of words for ‘language’ is in part to explore what these lexical items can tell us about how different societies conceptualize language. The starting point of Radden’s metonymic chain in Table 4.1 reflects the major significance of physiology and the human body in ‘folk’ understandings of language. That everyday notions of language so frequently find their conceptual source in body parts like ‘mouth’ and ‘tongue’ aligns in interesting ways with more scholarly conceptual frameworks that view language as an embodied phenomenon (e.g., Streeck et al., 2011). Another relevant dimension of many linguistic concepts of language is its communicative functions. We documented numerous lexical items that refer both to ‘language’ in general and to specific types of communication achieved through language, e.g., ‘news’, ‘discussion’, ‘issue’, ‘question’ and ‘quarrel’. The many different polysemous meanings of words for ‘language’ are discussed further in Section 3. Though we highlight certain conceptual patterns across languages, we acknowledge that our dictionary-based methods allow for only a superficial examination of concepts of ‘language’. Dictionaries offer a limited source of knowledge about the meanings of words, and this is especially true of the small-scale dictionaries that exist for most African languages, where in some cases Eurocentric biases have likely also crept in. The dictionaries we consulted all rely on a major European language of translation, which may of course interfere with how word meanings are represented. While our cross-linguistic investigation makes coarse-grained comparison possible, it would be usefully supplemented with more in-depth studies of language about language.³

3. As an interesting case in point, Irvine and Gunner (2018) use lexical evidence to argue that Zulu language ideologies are deeply rooted in bodily practice. In one part of their analysis, they

The chapter is organized as follows: in Section 2 we explain how we collected and coded our data and describe the contents of the database; in Section 3 we analyze the cross-linguistic data; in Section 4 we consider the unique polysemy ‘ear; language’ in Datooga in more detail; and in Section 5 we summarize our findings.

2. Data collection and coding strategies

To collect data for our cross-linguistic comparison of words for ‘language’, we consulted a wide range of online and printed dictionaries. We occasionally consulted language experts, but the majority of data comes from published sources. Our sample was one of convenience: we used online dictionaries published in the open-access journal *Dictionaria*, online vocabularies in the World Loanword Database, established online dictionaries (such as the online Chichewa dictionary),⁴ as well as printed dictionaries and grammars with vocabularies available in the library of the Institut für Afrikanistik at the University of Cologne. We restricted our sample to works published after 1950 and largely also to bidirectional dictionaries, where the language of translation was English, German, or French. However, we did consult monodirectional dictionaries when there was another dictionary available in the opposite direction. Our sample includes one language (Karimojong) for which we only had access to a monodirectional dictionary because we were able to quickly establish at least one polysemous meaning of the forms given for ‘language’ (by looking up body part terms).

In total we collected 174 lexical items glossed as ‘language’ from 130 languages spoken on the African continent (including several islands). This sample represents around 5.5% of languages currently spoken in Africa (using Glottolog’s estimate of 2,348 African languages; Hammarström et al., 2020). Since our data collection was guided entirely by the materials available to us, the sample is neither genetically nor geographically balanced. Using the top-level classification for each language as provided in Glottolog, our sample includes 50 Atlantic-Congo languages (38%), 40 Afro-Asiatic languages (31%), 8 Nilotic languages (6%), 8 Central Sudanic languages (6%), 7 Mande languages (5%), 4 Khoe-Kwadi languages (3%), and 13 others (including one creole). See our online Appendix (Mitchell and Zimmermann 2021) for a complete list of languages and data

discuss how metalinguistic descriptions of different varieties of Zulu make reference to certain tongue positions. Not only does the tongue (*u(lu)limi*) provide the conceptual source for the general word for ‘language’, but descriptions of configurations of the tongue help to distinguish varieties of Zulu.

4. <https://translate.chichewadictionary.org>, accessed 18 January 2021.

sources (link provided at end of paper). In terms of geographical distribution, using a rough classification of eastern, western, central, southern, and northern Africa, our sample looks as follows: 40% of the languages are spoken in East Africa; 33% in West Africa; 16% in southern Africa; 9% in Central Africa; and 2% in northern Africa. The assumed genetic and areal biases in our opportunistic sample mean we cannot make any robust claims about words for 'language' across Africa. Nonetheless, we are able to point to interesting variation in possible source concepts for 'language', including within language families.

When consulting bidirectional dictionaries, we looked up the word *language* (or equivalent) and made a note of all the forms given in the entry. (If there was no entry for 'language', we omitted this language from the database.) We then looked up each form in the opposite direction and made a note of all listed translations. One of our criteria for the inclusion of a given form in our database was that it was listed in the entry for 'language' and that 'language' was given as a meaning in its own entry. The purpose of this criterion was to ensure that 'language' was a stable and salient lexicalized meaning of a particular form and we discarded several lexical items on this basis.⁵ For each word for 'language' there were two possible scenarios: either the form was translated only as 'language' or it was presented as polysemous with one or more additional meanings. Our first broad coding category for each lexical item was therefore simply 'monosemous' or 'polysemous'.

For forms coded as 'monosemous', we made a note of any available information pertaining to etymology or morphologically related words. Such information was sometimes provided directly in the entry; in other cases, adjacent entries were clearly morphologically and semantically related. We also noted potential borrowings from other languages based on our existing knowledge or judgement, e.g., Swahili *lughā* and Tarifyt Berber *lluya* are Arabic borrowings. The etymological sources of apparently monosemous words for 'language' are highly relevant to an investigation of how words for 'language' develop, and Radden (2001) considered both polysemy and etymology in his study. However, lacking the necessary linguistic expertise for the vast majority of languages in the sample, our investigation of etymological relations was opportunistic and non-exhaustive. We comment on some diachronic sources for words for 'language' in Section 3, but the majority of the paper focuses on synchronic patterns of polysemy.

5. For example, in Alio und Jungraithmayr's (1989) dictionary of Bidiya (Chadic; Chad), we find the form *gārāw* under 'langue', but in the other direction, *gārāw* is translated only as 'parole'. We therefore excluded this particular lexical item. We made occasional exceptions to this 'two-way' criterion when additional evidence pointed to the robustness of the definition 'language'; e.g., we found the form *okwu* listed under 'language', but not vice versa, in Echeruo's (1998) Igbo dictionary, but the inclusion of the phrase *okwu bekee* 'English language' strongly suggests that *okwu* does have the meaning we are interested in.

For forms coded as ‘polysemous,’ we generally did not look for morphologically related words, even though a lexical item can obviously be synchronically polysemous and historically derived from semantically related forms. For example, the Datooga word for ‘language’ is *ítta*, which we coded as polysemous, with the additional meanings ‘ear’ and ‘news.’ This form may also be historically related to the verb *ii* ‘hear,’ but we did not typically comment on such relationships in our database. Again, this choice was partly based on insufficient linguistic knowledge of most of the languages in our database, but it also helped to constrain the scope of our study. Limiting ourselves to recording polysemous meanings was not always entirely straightforward either, however, on account of uncertainties about the translations provided for lexical items. One frequent problem we encountered was distinguishing the polysemous meanings of *tongue* in English and especially *langue* in French. Entries often distinguished between meanings with notes like ‘body part’ or ‘*organe*’ (for French) but sometimes we had to do additional research to be sure whether a form really meant ‘system of communication’ as well as ‘tongue (body part)’. We omitted forms from our database for which the nature of the polysemy was ambiguous. In a few cases, we also encountered heterosemy (same word form, different word class). All but one such cases also had additional meanings associated with the nominal form (i.e., were also polysemous), but we made a note of heterosemy as well.

Once we had collected over a hundred words for ‘language’ in different languages, we made a list of all the additional meanings that we had recorded. Based on our initial impression of the general frequency of each meaning, as well as the semantic domains discussed in Radden (2001), we came up with the semantic categories and subcategories given in Table 4.2 (and discussed further in Section 3). We then coded each lexical item in the database according to these categories. Depending on the number of senses given in the dictionary entry, lexical items could be grouped under multiple categories. The highest number of meaning categories associated with a single form was five. Both authors went through all the data and agreed on the coding.

In a few cases, the other meanings given for a lexical item glossed as ‘language’ were not obviously polysemous with ‘language.’ For example, in Rongier’s (1995) Ewe dictionary, the noun *gbè* has the following senses (translated from French): (i) day, date; (ii) voice, sound, noise; (iii) language; (iv) time. We coded the form *gbè* as polysemous (and categorized it under BODY PART: VOICE as well as OTHER for ‘sound’) because we perceive a semantic relation between senses (ii) and (iii), but the relation between senses (i), (iii), and (iv) is much less apparent, at least to non-Ewe speakers. In another Gbe language in our sample, Fon, there are two separate entries for *gbè*, one glossed as ‘language’ and the other as ‘day,’ suggesting that these forms might be homophonous rather than polysemous in Ewe as well.

Table 4.2 Categories for coding polysemy in words for ‘language’

Broad semantic category	Subcategories
A. Body part	1. Tongue 2. Mouth 3. Throat 4. Voice 5. Ear 6. Lip
B. Linguistic unit	1. Word 2. Sentence
C. Linguistic action (e.g., ‘speech’)	
D. Genre of communication (e.g., ‘news’, ‘matter’, ‘information’)	
E. Ethnic group	
F. Other	

We decided not to include ‘day’ as a polysemous meaning, though this may turn out to be incorrect. Such uncertainties were rare, however.

Once we had collected and coded data from as many dictionaries as we could conveniently consult, we were able to explore cross-linguistic patterns in words for ‘language’.

3. Cross-linguistic findings

As noted above, each word for ‘language’ in our database was initially classified as either monosemous or polysemous. This section deals briefly with possible sources of those lexical items coded as monosemous and then considers cross-linguistic patterns of polysemy. We first discuss the broad semantic categories we used to classify polysemous meanings and then consider their relative frequency, drawing comparisons with Radden’s (2001) study. We also comment on the geographical distribution of some common patterns of polysemy, though our convenience sample is not adequate for making genetic or areal claims. Another limitation of our methods, and thus our results, is our dependence on dictionaries as a source of information about word meanings.

46 (26%) of the words we collected for ‘language’ had no other senses listed in their dictionary entry and were coded as monosemous.⁶ We identified 11 of the monosemous forms as borrowings, with either Arabic or Amharic as the donor language. Another 14 forms were identified (to varying degrees of certainty) as derived from or otherwise morphologically related to another lexical item, in almost all cases a verb meaning ‘speak’ or ‘say’ or a body part metonymically related to speaking. As explained above, we did not include these semantic relations in our main analysis, but these examples provide additional support for the claim that words for ‘language’ commonly develop from concepts of speaking as an activity and from body parts associated with speaking. A small number of monosemous words in our database appear to be formed by means of compounding. All but one of these forms come from Mande languages and combine a noun meaning ‘kind, sort; race, nation; descendent’ with a noun meaning ‘word; matter; problem’, e.g., *buiyā* ‘language’ in Boo (Mande; Benin, Nigeria) from *bui* ‘kind, sort, species, race, nation, descendant’ and *yā* ‘word, matter, affair, account, problem, case, verdict’ (Jones, 2010). Another binominal compound in our database comes from Degema (Benue-Congo; Nigeria): *enu-ekeny* ‘language’, from *enu* ‘mouth, burrow, opening’ and *ekeny* ‘town, village, city, community’ (Kari, 2008).⁷ In both cases, one part of the compound denotes a concept that we commonly encounter with polysemous words for ‘language’ as well – ‘word’ and ‘mouth’, respectively – where this concept is combined with a noun denoting geographic or ethnic origin. Finally, 14 of the monosemous forms in our database have no accompanying notes as to their possible source. It is conceivable that some of these lexical items therefore constitute basic words for ‘language’, though such words are predicted to occur rarely (Radden, 2001, p. 55). It seems more likely that experts in the respective languages would be able to identify lexical relationships that are inaccessible to us based on dictionaries alone.

Turning to the 128 polysemous words for ‘language’ in the database, we first discuss the semantic categories we used for coding the data and then consider the frequency of these different categories. Radden (2001, p. 55) identifies the following broad semantic domains which encompass many polysemous meanings of words for ‘language’: “(i) articulation and speech organs, (ii) linguistic action, and (iii) basic linguistic units”. With respect to (i), our ‘speech organ’ (or ‘body part’) categories (see Table 4.2) are based simply on what appeared in the data,

6. One word for ‘language’ in Wolof had additional senses but these were for a heterosemous verb form.

7. One further language, Tera (Chadic; Nigeria), has a very similar ‘mouth’ + ‘town’ compound form, *me ghai*, but in addition to ‘language’ this form is listed with the meanings ‘tradition’, ‘custom’, ‘ethnic group’ and was categorized as polysemous.

regardless of frequency. The meanings associated with these categories are fairly self-explanatory since they refer to concrete, relatively discrete parts of the body, though the semantic extension of ‘throat’ is not always easy to delineate – we included the single occurrence of French *cou* ‘neck’ in this category.⁸ ‘Voice’ is of course not a physical body part, but we follow Radden (2001, p. 57) in treating it as a speech organ. Radden (2001, p. 63) identified the following speech organs as “metonymic vehicle[s] for ‘speaking’, ‘speech’ and ‘language’”: ‘voice’, where ‘voice’ is metonymically related to ‘throat’, ‘tongue’, ‘mouth’, and ‘lips’. We found all of these polysemous meanings in our data, plus ‘ear’, a novel body part not mentioned by Radden. The metonymy of ‘ear’ for ‘language’ is discussed in detail in Section 4. The absence noted in Radden (2001, p. 74) of the metonymy TEETH FOR LANGUAGE was confirmed in our study.

Radden’s semantic domain ‘linguistic action’ was strongly apparent in our data, too. In this category we include meanings such as ‘speech’, ‘talk’, ‘talking’, and ‘telling’. With respect to ‘basic linguistic units’, we found ‘word’ and ‘sentence’ as polysemous meanings of words for ‘language’. Two semantic domains not mentioned by Radden that occurred in our sample are ‘genre of communication’ and ‘ethnic group’. The latter category includes meanings translated as ‘tribe’ and ‘*groupe ethnique*’. For instance, Ihanzu (Bantu; Tanzania) has a word *ntambu*, reported to mean both ‘language’ and ‘ethnic group’ (Andrew Harvey, p.c.). This polysemy hints at possible ideological links between language and identity, though any conceptual analysis of the few words in this category would require much more detailed information about their meaning and use. ‘Genre of communication’ is our most nebulous category, intended to account for the many examples where a word for ‘language’ also refers to a metapragmatic category or type of linguistic communication, such as ‘news’, ‘matter’, ‘affair’, ‘issue’, ‘discourse’, ‘problem’. Although Radden does not discuss such meanings, ‘genre of communication’ can be accounted for in his scheme (see Table 4.1) by means of the metonymy ACTION FOR RESULT, whereby a word for ‘speaking’ comes to be used for metapragmatic “objects” or genres of speech. In some cases, these metapragmatic genres may then extend beyond the verbally articulated to give rise to more general meanings such as ‘issue’ and ‘problem’. Finally, our ‘other’ category includes a wide range of word meanings that were not easily subsumed under one of the other categories. We take a closer look at the contents of this category at the end of this section.

Table 4.3 presents the overall frequency of each of the polysemous meaning categories that we established in Table 4.2. As already mentioned, a given lexical item could appear in multiple semantic categories depending on its range of poly-

8. The definition for the same form also included *gorge* ‘throat’.

semous meanings. For example, in the Geez language of Ethiopia, the form *nagar* is translated as ‘language’, ‘speech, talk’ (LINGUISTIC ACTION), ‘word’ (LINGUISTIC UNIT), and ‘subject, matter, discourse’ (GENRE OF COMMUNICATION) and was included in each of these four categories. On the basis of our coding scheme, the 128 polysemous words were sorted into categories 236 times in total. The percentages given in Table 4.3 give us an impression of the relative frequency of each polysemy, based on all the different meanings associated with the 128 polysemous words for ‘language’. The largest number of meanings additional to ‘language’, just under a quarter, were assigned to our ‘other’ category, discussed below. Almost a third of meanings fall under one of the ‘body part’ subcategories, with ‘mouth’ being the most frequent source concept for ‘language’ in our sample. This finding contrasts to some extent with Radden (2001, p.80), who found the metonymy MOUTH FOR LANGUAGE to occur only “occasionally”. In order of frequency, we then find ‘tongue’, ‘voice’, ‘lip’, ‘throat’ and ‘ear’ as polysemous meanings of words for ‘language’. The rarity of ‘lip’ in our data aligns with Radden’s observations. The next most frequent broad semantic domain associated with words for ‘language’ is LINGUISTIC ACTION (‘SPEECH’). Radden (2001) identified several world languages where the word ‘word’ also means ‘language’; this polysemy was observed in over a tenth of our lexical items. In contrast, ‘sentence’ occurred only in a single language as an additional meaning of ‘language’ (in the Bantu language Talinga-Bwisi). Finally, our own categories of ‘genre of communication’ and ‘ethnic group’ were relatively infrequent as well.

Where in Africa do we find these different polysemous relations? We cannot make any robust generalizations about areal semantic patterns on the basis of our convenience sample, but we can at least give some impression of the geographical distribution of source concepts for ‘language’ in the data we collected. Map 4.1 consists of four panels, each of which shows the approximate geographical location of language communities for which a word for ‘language’ also means (a) ‘mouth’, (b) ‘tongue’, and (c) ‘speech’, while (d) maps languages in which a word for ‘language’ is monosemous.⁹ Note that a language that appears on one of these maps may have multiple words for ‘language’, and the words themselves

9. The maps were created in R (R Core Team, 2020) using the package *lingtypology* (Moroz 2017), which connects R with the Glottolog database (Hammarström et al., 2020). In order to use this package to create the maps, we had to choose language names as given in Glottolog, even if they differed from the names used in the dictionaries that were consulted for our database. In addition, we had to use the level of genetic classification for which geographical coordinates are available in Glottolog. As a result, it was not always possible to choose the specific language variety named in a given dictionary, because we had to use the next highest classification. The geographical coordinates themselves should of course be understood as approximate: many of the languages in our sample are spoken by communities across multiple regions.

Table 4.3 Polysemous meanings of words for ‘language’ organized by frequency

Semantic category	Number of words for ‘language’ glossed with this meaning	%
Other	56	23.7%
Linguistic action	40	17.0%
Body part: mouth	34	14.4%
Body part: tongue	30	12.7%
Linguistic unit: word	29	12.3%
Genre of communication	23	9.7%
Body part: voice	11	4.6%
Ethnic group	6	2.5%
Body part: lip	3	1.3%
Body part: throat	2	0.9%
Body part: ear	1	0.4%
Linguistic unit: sentence	1	0.4%
Total	236	

could have multiple polysemous meanings, which means that a language could appear on multiple maps. The most striking feature of Map 4.1 is that languages with the polysemy ‘mouth; language’ are clustered in eastern and west-central Africa, with no cases appearing outside of these regions, at least in our sample. This polysemy, represented by 34 lexical items in our database, is found in many different language families (Nilotic, Afro-Asiatic, Ta-Ne-Omotiic, Atlantic-Congo, and Central Sudanic) and has conceivably been transmitted horizontally in some cases. In contrast, the other two most common patterns of polysemy are distributed fairly widely across the continent. Languages in our sample with a monosemous word for ‘language’ are also widely dispersed across Africa. (Recall that this classification does not necessarily mean that the word ‘language’ is basic in these languages – see discussion of borrowings and compounds above.)

Lastly, we come to the contents of our OTHER category. This category accounted for almost a quarter of the polysemous meanings of words for ‘language’, but the majority of these senses extend from a more basic meaning of a given lexical item, rather than relating directly to ‘language’. For instance, senses categorized under OTHER like ‘blade’, ‘flame’, and ‘tip’, or ‘opening’, ‘entrance’, and ‘doorway’, are obviously extensions from ‘tongue’ and ‘mouth’, respectively, and thus relate to ‘language’ only indirectly. Similarly, we noted the polysemous senses ‘noise’ and ‘sound’ in several cases, but these meanings most likely extend from the more basic (and always co-occurring) meaning ‘voice’. Another sense occa-

4. ‘Language’ in Datooga

Of all the languages in our sample, the Nilotic language Datooga is the only one to exhibit the polysemy ‘language’ and ‘ear’. This section explores this polysemy further with data collected by the first author from the Gisamjanga and Barabaiga dialects of Datooga, as spoken in communities living in Manyara Region, Tanzania. Our analysis is based only on linguistic data that was collected with other projects in mind; metalinguistic discussions with speakers would contribute considerably to the interpretations presented here. In addition to ‘ear’ and ‘language’, the word *íttá* can also mean ‘news, information’. Examples (1)–(3), taken from recordings of spontaneous conversation in Datooga-speaking households, exemplify the various meanings. Example (1) illustrates the use of *íttá* to refer to the body part ‘ear’. This utterance comes from a conversation about a deaf woman and the extent of her linguistic knowledge. The speaker makes the point that the woman’s disability is purely anatomical rather than cognitive – her ear is ‘blocked up’:¹⁰

- (1) *mêaqús gídêabà géedùkt íttá sída*
m-êa-qús gídêabà g-ée-dùg-d-á íid-dá síi-da
 NEG-IMPS-be.only CONJ AFF-IMPS-block-CF-IS ear-SG.POSS person-SG
 ‘Isn’t it just that the person’s ear is blocked up?’

In Example (2), a child asks a metalinguistic question about how to translate a Swahili expression into his native language, using the conventional possessive phrase *íttá Dàtóogá* to refer to the Datooga language:

- (2) *àbà íttá Dàtóogá náahá*
àbà íid-dá Dàtóogá náa=há
 PREP ear-SG.POSS Datooga what=DSC
 ‘What is it [the number 2,000] in the Datooga language?’

The form *íttá* is used to refer to other languages, too, e.g., *íttá nàyéegída* ‘Iraqw language’.

10. We use the concrete, basic meaning of *íttá*, namely ‘ear’, as the morphological gloss for this lexical item to make the polysemy clear, though the extent to which speakers understand these different meanings as related is not known. Abbreviations used in the morphological glosses are those defined in the Leipzig Glossing Rules, with the following additions: AFF = ‘affirmative’; CF = ‘centrifugal’; CONJ = ‘conjunction’; DSC = ‘discourse marker’; IMPS = ‘impersonal’; IS = ‘inflectional suffix’; PREP = ‘preposition’. We use a slightly adapted version of the orthography developed by the Datooga Bible Translation Project.

Finally, (3) demonstrates the ‘news, information’ sense of *íttá*. Here, a woman asks her adult son whether he has heard the news from home (the announcement of a *wérwérge* celebration for a young child):

- (3) *nínyás íttá qêeda?*
n-í-nyás-í íid-dá qée-da
 PRF-2SG-get-IS ear-SG.POSS house-SG
 ‘Have you heard the news from home?’

All three meanings are relatively well attested in the available corpus data. The plural form, *íidiga*, is associated with the same polysemous meanings, i.e., ‘ears’, ‘languages’, and ‘news, information’.

Additional evidence for a conceptual association between ‘ear’ and ‘language’ comes from the Datooga avoidance vocabulary. Other studies have drawn on respect registers to explore semantic relations, e.g., Dixon (1971) or Evans and Wilkins (2000) on the semantic extensions of perception verbs in Australian languages.¹¹ In the Datooga in-law name avoidance register (see Mitchell (2015) for details), we generally find a one-to-one mapping between ordinary words and their avoidance counterparts, though in the case of polysemous words, we sometimes observe a one-to-many relationship. For instance, the everyday noun *húlánda* can mean (i) ‘men’s house’, (ii) ‘clan meeting’, or (iii) ‘youth meeting’. Meaning (ii) likely developed from (i) through a process of metonymy, since clan meetings may take place in the physical space of the men’s house. In the avoidance register, each of these three senses is expressed differently. This one-to-many lexical relationship possibly reflects that the meanings are not perceived as closely related to one another by speakers. In other cases, patterns of polysemy are carried over into the avoidance register. With respect to *íttá*, the avoidance counterpart shares the polysemous structure of the everyday form: the word *mínista* (a deadjectival noun from *mínis* ‘deaf’) replaces *íttá* and can mean ‘ear’, ‘language’ or ‘news’.¹² For example, *íttá Nàyéegìida* ‘Iraqw language’ is rendered as *mínistá Gíttínóodiga* in the avoidance register. Thus the conceptual link between auditory

11. Australian respect registers (or ‘in-law languages’) often exhibit one-to-many lexical relationships with the everyday vocabulary, i.e., a single lexical item in the respect register corresponds to multiple lexical items in the everyday register (Dixon, 1971). In their exploration of the extension of verbs of hearing to verbs of knowing, Evans and Wilkins (2000) found that the ordinary Kuwinjku/Mayali language has distinct forms for ‘hear’ and ‘know’, but that the respect register in this language conflates the meanings ‘listen, hear’ and ‘understand, know’ into a single form, providing another kind of evidence for this conceptual link across Australia.

12. Several alternative avoidance forms for ‘ear’ have also been documented, including *mínischêanda* (from the same root *mínis* ‘deaf’) and *dáwshìitta* (possibly related to *dàwìishádèeda* ‘molar; lower jaw’).

perception and language is revealed in the avoidance register as well as the every-day language.

Another kind of evidence for the link between auditory perception and language can be observed in the expression ‘to hear a language’, meaning ‘to know a language’, as illustrated in (4):

- (4) *màýínyí úttá Dátóogá?*
m-à-úny-i úid-dá Dátóo-gá
 NEG-3-hear-IS ear-SG Datooga-PL
 ‘Doesn’t s/he speak Datooga?’

In contrast with our English translation, the Datooga expression prioritizes hearing over speaking in conceptualizing linguistic ability.¹³

To reiterate, use of the source concept ‘ear’ for ‘language’ is not mentioned in Radden (2001) and is not attested for any other language in our sample. This novel semantic pattern can nonetheless be largely accounted for in terms of the metonymic processes already known to give rise to abstract concepts of ‘language’. Instead of a metonymic pathway leading from speech articulators such as ‘mouth’ or ‘tongue’ to the act of speaking, here the metonymy begins with an organ of auditory perception. Accordingly, the next conceptual link in the metonymic chain is not speaking, but hearing, though the metonymy INSTRUMENT FOR ACTION captures both kinds of extension. The act of ‘hearing’ is not a lexicalized meaning of *úttá* – this is an example of an “intermediate stage” that gets skipped in the metonymic chain (Radden 2001, p.59) – but the function of the ear clearly plays an important role in the conceptual extension of *úttá*.¹⁴ (Semantic extension of the body part *úttá* to the activity or function associated with it is found in certain expressions, such as the verbal phrase *leaysiin úttá* ‘listen’, literally ‘drink [something] into the ear’.) The metonymy ACTION FOR RESULT discussed by Radden (2001) takes us from the concept of ‘hearing’ to the more abstract idea of what is heard, which then may be conceptualized more specifically as ‘news’ or ‘information’.¹⁵ That is, the ‘result’ of auditory perception via the ear is a recognizable genre or type of communication. More abstractly still, the results or ‘objects’ of auditory perception can also be conceptualized in terms of the manner or system of communication being employed. This semantic exten-

13. An anonymous reviewer suggests this is common across Africa. A cross-linguistic project on expressions relating to language competence would be an interesting project in itself.

14. The historical direction of a metonymy ‘EAR FOR HEARING’ is in fact uncertain in Datooga, since the verb ‘hear’ in Datooga is *ii(ny)*, from which the noun *úttá* possibly derives.

15. The English language lacks a nominalization of the verb *hear* equivalent to *speak* → *speech*, hence we use the slightly awkward ‘what is heard’.

sion of ‘ear’ relates not to *what* is communicated and heard (as in ‘news, information’) but rather *how*, i.e., in what medium (as in ‘language’). This interpretation of the relationship between the three senses of *ítta* is non-linear, with both ‘news’ and ‘language’ extending independently from ‘ear’ via the intermediate metonymy HEARING FOR WHAT IS HEARD. An alternative proposal is that the ‘language’ sense developed out of the ‘type/object of communication’ (i.e., ‘news’) sense. Aligning with Radden’s model, we could treat ‘what is heard’ as conceptually parallel to ‘speech’ and explain the extension to ‘language’ in the same way, invoking the SPECIFIC FOR GENERIC metonymy (i.e., ‘what is heard’ “is a specific instantiation of language in general” (Radden, 2001, p.58).¹⁶ Though this account of the link between ‘what is heard’ and ‘language’ no doubt oversimplifies the historical conceptual trajectory, it allows us to accommodate the unusual Datooga pattern within a slightly modified version of Radden’s existing cross-cultural schema, as indicated in Table 4.4:

Table 4.4 Modified version of Table 4.1 that accommodates ‘ear for language’ metonymy

speech organ	> speaking /hearing	> what is spoken/heard	> language
specific metonymy	SPEECH ORGAN FOR SPEAKING / HEARING	SPEAKING / HEARING FOR WHAT IS SPOKEN / HEARD	WHAT IS SPOKEN / HEARD FOR LANGUAGE
general metonymy	INSTRUMENT FOR ACTION	ACTION FOR RESULT	SPECIFIC FOR GENERIC

The EAR FOR LANGUAGE metonymy evidenced in Datooga appears to be an innovation not shared by other Nilotic languages. We have eight Nilotic languages in our sample, with 2–3 languages represented from each subfamily (using Glottolog labels). Table 4.5 presents the lexical items we collected for these languages and their associated meanings (see online Appendix for references).

Based on this small amount of data, we observe that the speech organs ‘mouth’ and ‘tongue’ appear to be the main sources for the concept ‘language’ in Nilotic. Interestingly, the Eastern Nilotic languages Turkana and Karamojong have two words for ‘language’, with both of these body parts as source concepts. In the Southern Nilotic family, to which Datooga belongs, we find both ‘mouth’ and ‘tongue’ extended to mean ‘language’, though this pattern does not obtain

16. Alternatively, the link from ‘what is heard’ to ‘language’ could potentially be explained in terms of the general metonymy OBJECT FOR MATERIAL CONSTITUTING THE OBJECT (Radden & Kövecses, 1999, p.32), if ‘language’ were construed as a kind of material or substance.

Table 4.5 Words for ‘language’ in Nilotic languages

Language	Subfamily	Word for ‘language’	Meanings
Luo	Western Nilotic	<i>dhó</i>	‘mouth; language’
Anuak	Western Nilotic	<i>dhók</i>	‘mouth; language’
Turkana	Eastern Nilotic	<i>akituk</i>	‘mouth; language’
		<i>angajep</i>	‘tongue; language’
Teso	Eastern Nilotic	<i>anajep</i>	‘tongue; language’
Karamojong	Eastern Nilotic	<i>anajep</i>	‘tongue; language’
		<i>akituk</i>	‘mouth; language’
Nandi	Southern Nilotic	<i>kutit</i>	‘mouth; language’
Markweeta	Southern Nilotic	<i>ngelep</i>	‘tongue; language’
Barabayiiga-Gisamjanga	Southern Nilotic	<i>íttá</i>	‘ear; language’

for Datooga (Barabayiiga-Gisamjanga). In Datooga, *qúuta* ‘mouth’ (cognate with Nandi *kutit*) carries various polysemous meanings, including ‘opening’, ‘beside’, ‘top’, as well as ‘voice’, but does not conventionally extend to mean ‘language’. Conversely, in the Nandi dictionary, *itit* ‘ear’ is listed with this single concrete meaning. However, the entry does include the expression ‘to hear by ear’, where ‘ear’ contributes the meaning ‘for oneself’, suggesting a possible extension from the literal body part to the more abstract idea of evidential source. Still, this in no way resembles the polysemous meanings of Datooga *íttá*. While the morpheme *i:t ‘ear’ has been reconstructed for Proto-Southern Nilotic (Rottland, 1982), its semantic trajectory appears to have taken a rather different course in Datooga. We have reconstructed this trajectory in Table 4.4, though why ‘hearing’ played a more significant role than ‘speaking’ for Datooga in the development of their word for ‘language’ remains an outstanding question. Based on the lexical data alone, we might conjecture that Datooga speakers have historically placed greater value on listening than speaking in their understandings of communication. Such speculation could be probed in future through detailed ethnographic work and metalinguistic discussion with Datooga people.

5. Conclusion

We collected words for ‘language’ and their additional senses in 130 African languages with the goal of identifying patterns in the source concepts for this abstract notion. Around three quarters of the words we collected were polyse-

mous. As Radden (2001) also demonstrated, the meaning ‘language’ typically extends from more basic semantic concepts. In our data, source domains for ‘language’ included body parts involved in language production and comprehension, linguistic activities (‘speaking’, ‘talking’), and genres or “content” of communication (‘news’, ‘discourse’). Other patterns of polysemy, where the directionality was less clear, included lexical items that mean both ‘language’ and ‘ethnic group’ as well as ‘language’ and ‘word’. The body parts that we identified as providing source concepts for ‘language’ were consistent with Radden’s (2001) findings, with the exception of *EAR FOR LANGUAGE*, a metonymy found so far only in Datooga (Nilotic; Tanzania). In addition, while Radden highlighted ‘tongue’ as an especially salient speech articulator and source concept for ‘language’, we found ‘mouth’ to be slightly more frequent than ‘tongue’ as a polysemous meaning of ‘language’ in African languages. However, as shown in Map 4.1, *MOUTH FOR LANGUAGE* is geographically more restricted than *TONGUE FOR LANGUAGE*.

Our exploration of polysemy in words for ‘language’ was motivated by an interest in how African communities have conceptualized linguistic phenomena. Lexical patterns as gleaned from dictionaries can only tell us so much, of course, and in-depth, ethnographic studies of language ideologies and the “natures of language” (Hauck & Heurich, 2018) are indispensable here. Nonetheless, an assumption at the heart of much research in anthropological linguistics is that word meanings emerge out of habitual, culturally embedded interactions and countless situated acts of interpretation, such that a metonymic pattern like *MOUTH FOR LANGUAGE* points to certain histories of thinking. One basic observation we can make from our data is that objectified notions of ‘language’, as reflected in the existence of nouns with this meaning, are widespread and presumably serve useful functions in everyday social life (as well as in scholarly discourse). Yet we also observe that these everyday concepts are typically rooted in ideas about action and process in the material world. Returning to Radden’s (2001) folk model of language, his preliminary suggestion that these folk models ultimately develop from concepts relating to speech articulation turns out not to hold, at least not universally. However, if we conceive of speech and speech organs more broadly to include the articulators as well as receptors, we can preserve the claim that people’s words for language have a strong basis in physical, embodied aspects of communication. The ways humans think about language and linguistic behavior, as revealed through their words for ‘language’, support a theory of meaning as “grounded in our sensorimotor experience” (Johnson & Lakoff, 2002, p.245). Nonetheless, within the bounds of these shared processes of meaning-making, we do observe some degree of cultural specificity in the aspects of linguistic experience drawn on to make abstract reference to ‘language’, whereby mouths, tongues, and ears are all possible source concepts.

References

- Alio, K., & Jungrathmayr, H. (1989). *Lexique bidiya: Une langue centre-africaine (République du Tchad)*. V. Klostermann.
-  Bagasheva, A. (2017). Cultural conceptualisations of mouth, lips, tongue and teeth in Bulgarian and English. In F. Sharifian (Ed.), *Advances in cultural linguistics* (pp. 189–221). Springer.
- Dixon, R. M. W. (1971). A method of semantic description. In D. D. Steinberg & L. A. Jakobovits (Eds.), *Semantics. An interdisciplinary reader in philosophy, linguistics, and psychology* (pp. 436–471). Cambridge University Press.
- Doke, C. M., D. M. Malcom, J. M. A. Sikakana, & B. W. Vilakazi. (1990). *English Zulu–Zulu English dictionary*. Witwatersrand University Press.
- Echeruo, M. J. C. (1998). *Igbo-English dictionary. A comprehensive dictionary of the Igbo language, with an English-Igbo index*. Yale University Press.
-  Evans, N., & Wilkins, D. (2000). In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language*, 76(3), 546–592.
-  Goddard, C. (2011). The lexical semantics of language (with special reference to words). *Language Sciences*, 33(1), 40–57.
- Goddard, C., & Wierzbicka, A. (2014). *Words and meanings. Lexical semantics across domains, languages, and cultures*. Oxford University Press.
-  Goosens, L. (2009). Metaphonymy: The interaction of metaphor and metonymy in expressions for linguistic action. *Cognitive Linguistics*, 1(3), 323–342.
-  Hammarström, H., Forkel, R., Haspelmath, M., & Bank, S. (2020). *Glottolog*, 4(3). Max Planck Institute for the Science of Human History.
-  Hauck, J. D., & Heurich, G. (2018). Language in the Amerindian imagination: An inquiry into linguistic natures. *Language & Communication*, 63, 1–8.
- Irvine, J. T., & Gal, S. (2000). Language ideology and linguistic differentiation. In P. V. Kroskrity (Ed.), *Regimes of language; ideologies, politics, and identities* (pp. 35–83). School of American Research Press.
-  Irvine, J. T., & Gunner, L. (2018). With respect to Zulu: Revisiting ukuHlonipha. *Anthropological Quarterly*, 91(1), 173–207.
-  Jing-Schmidt, Z. (2008). Much mouth much tongue: Chinese metonymies and metaphors of verbal behaviour. *Cognitive Linguistics*, 19(2), 241–282.
-  Johnson, M., & Lakoff, G. (2002). Why cognitive linguistics requires embodied realism. *Cognitive Linguistics*, 13(3), 245–263.
- Jones, R. M. (2010). *Dictionnaire Boo-Français-Anglais: Avec index Français-Boo*. Lincom.
- Kari, E. E. (2008). *Degema-English dictionary with English index*. Research Institute for Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies.
-  Lüpke, F., & Storch, A. (2013). *Repertoires and choices in African languages*. De Gruyter Mouton.
- Makoni, S., & Mashiri, P. (2007). Critical historiography: Does language planning in Africa need a construct of language as part of its theoretical apparatus? In S. Makoni & A. Pennycook (Eds.), *Disinventing and reconstituting languages* (pp. 62–89). Multilingual Matters.

-  Mitchell, A. (2015). Words that smell like Father-in-Law: A linguistic description of the Datooga avoidance register. *Anthropological Linguistics*, 57(2), 195–217.
- Moroz, G. (2017). lingtypology: Easy mapping for Linguistic Typology. <https://CRAN.R-project.org/package=lingtypology>
-  Newman, P., & Schuh, R. G. (2016). Hausa language names and ethnonyms. *Journal of African Languages & Linguistics*, 37(2), 185–200.
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved on 21 June 2023 from <https://www.R-project.org/>
- Radden, G. (2001). The folk model of language. *metaphorik.de*, 1, 55–86.
-  Radden, G. & Kövecses, Z. (1999). Towards a theory of metonymy. In K.-U. Panther & G. Radden (Eds.), *Metonymy in language and thought* (pp. 17–59). John Benjamins.
- Rongier, J. (1995). *Dictionnaire français-éwé. Suivi d'un index éwé-français*. A.C.C.T.
- Rottland, F. (1982). *Die südnilotischen Sprachen. Beschreibung, Vergleichung und Rekonstruktion*. D. Reimer.
- Streeck, J., Goodwin, C., & LeBaron, C. (Eds.). (2011). *Embodied interaction. Language and body in the material world*. Cambridge University Press.
-  Sweetser, E. E. (1992). English metaphors for language: Motivations, conventions, and creativity. *Poetics Today*, 13(4), 705–724.
-  Wei, L. (2018). Translanguaging as a practical theory of language. *Applied Linguistics*, 39(1), 9–30.

Appendix

-  Mitchell, Alice, & Zimmermann, Nicola. (2021). Source concepts for language data [Data set]. *Zenodo*.