LUNDS
UNIVERSITET

Tracking Linguistic Primitives

The Phonosemantic Realization of Fundamental Oppositional Pairs

Niklas Johansson

Centre for Language and Literature
Master's Thesis in General Linguistics
December 2013
Supervisor: Gerd Carling
Co-supervisors: Arthur Holmer, Joost van de Weijer, Jordan Zlatev

Abstract

This thesis investigates how cross-linguistic phoneme distributions of 56 fundamental oppositional concepts can reveal semantic relationships by looking into the linguistic forms of 75 genetically and areally distributed languages. Based on proposals of semantic primes (Goddard 2002), reduced Swadesh lists (Holman et al. 2008), presumed ultraconservative words (Pagel et.al. 2013), attested basic antonyms (Paradis, Willners & Jones 2009) and sense perception words, a number of semantic oppositional pairs were selected. Five different types of sound groupings were used dividing phonemes according to; the frequency of vowels' second formant and consonants' energy accumulation (Frequency), sonority (Sonority), a combination of the aformentioned two (Combination), general phonetic traits, e.g. voicing (General), and lastly incorporating all traits of the four presented groupings (All). These were analyzed by means of cluster analyses creating biplots, illustrating the phonological relatedness between the investigated concepts. Also, the phoneme distributions' over- and underrepresentation from the average was calculated defining which sounds represented and were lacking for each concept. Significant semantic groupings and relations based solely on phonological contrasts were found for most investigated concepts, including the semantic domains; Small, Intense Vision-Touch, Large, Organic, Horizontal-Vertical Distance, Deictic, Containment, Gender, Parent and Diurnal, and the sole concept OLD. The most notable relations found were; MOTHER/I vs. FATHER, a three-way deictic distinction between I, indicatory deictic concepts and THERE, and a dimensional tripartite oppositional relationship between Small and (possibly with Intense Vision-Touch), Large-Organic and Horizontal-Vertical Distance. Embodiment, benefits of oppositional thinking and evidence for more general concepts to precede complex concepts were proposed as explanations for the results.

Keywords: phonosemantics, sound symbolism, iconicity, non-arbitratiness, semantic typology, phonetic typology, universals

Acknowledgements

I would like to thank my supervisors Gerd Carling, Arthur Holmer, Joost van de Weijer and Jordan Zlatev for assistance and input during the writing of this thesis especially regarding discussions considering semantics, typology and statistical analyses. I would also like to thank Mechtild Tronnier for guidance concerning phonetics, Revaz Tchantouria for providing data of Georgian and Inga-Lill Hansson for providing data of Akha. It has all been very much appreciated and I am very grateful for your help.

Contents

1. Introduction

The phenomenon of *phonosemantics*, also called *sound symbolism*, *non-arbitrariness*, *iconicity* etc., has been subject for debate for more than two thousand years, both regarding its existence and in what domains it operates. As noticed already by Jespersen (1922), certain semantic domains are more likely to be phonosemantically motivated than others; in particular this concerns semantic qualities closely connected to sensory perception (e.g. size, shape, texture) or deixis. And following the abundance of experimental results during the 20th century and onward, most scholars now agree that phonosemantics is a linguistic factor which cannot be ignored and that every language contains lexemes that are phonosemantically motivated. Many investigations have however been focused on one sole semantic domain and usually only incorporating a few speech sounds, e.g. the rather famous experiments by Sapir (1929) regarding size and by Köhler (1929) regarding shape. Though, some recent investigations (e.g. Wichmann, Holman & Brown 2010) have shown that by using large samples of languages and very fundamental concepts it is possible to find associations, not only between sound and meaning, but also between different meanings after the form-meaning relation has been established. And further, motivated form-meaning associations are typically realized as either oppositional contrasts (e.g. deixis), gradient relations (e.g. color terms), or complex clusters (e.g. phonesthemes). What this indicates is that finding cross-linguistic phonosemantic relationships could allow us to get a glimpse of which semantic domains could constitute potential linguistic primitives and hence tell us something about the course of the evolution of language as a whole.

1.1 Research Questions

The primary goal of this investigation is to tie together many of the previous studies of phonosemantics. Oppositional word pairs of basic vocabulary seem to be heavily affected by phonosemantics, hence these are the object of study and will be thoroughly presented in the theoretical background. Furthermore, since closely related concepts tend to function in the same way, these might share at least some similar sound patterns. The secondary goal is to investigate which sounds represent which concepts, provided that cross-linguistic similar sound patterns are found. These patterns could in turn be relevant for indicating (perhaps

universal) cognitive mappings between concepts and sounds and between different concepts. Hence the research questions which read as follows:

1) Will oppositional word pairs of basic vocabulary show correlations between semantically related meanings solely based on their phonetic makeup?
2) If such correlations are found, what is the phonetic makeup of each concept?
3) If such correlations are found, which semantic domains and which relations between the semantic domains are found?

In order to answer these questions, statistical and quantitative methods were employed. In Chapter 2, the theoretical background is presented, beginning lexical universals in section 2.1, including studies of Swadesh Lists and other base vocabulary related research, as well as the role of oppositions in cognition. Section 2.2 provides an overview of phonosemantics along with experimental results yielded by various authors and a summary of which semantic domains usually are affected by phonosemantics, as well as what role phonetic dimensions play.

Chapter 3 describes the methods of the study; how a genetically and areally balanced sample of languages was created and the selection of potentially interesting and manageable concepts to investigate, sections 3.1 and 3.2. This chapter also includes quantification of data; how the phonemes' of the featured concepts were divided according to different phoneme groupings, section 3.3, and data sources, section 3.4. The analysis procedures, i.e. how the concepts' similarity among each other based on phonetic makeup was calculated using statistical computing and graphics software R and compared with semantic qualities, and how the phoneme distributions' deviation from the normal was calculated are presented in section 3.5.

Chapter 4 presents general findings; concepts judged significant for each phoneme grouping, the semantic domains found, based on the found concept clusters and their relations among each other. Further, sound makeup of each concept is also presented.

Chapter 5 offers a more general discussion of the results, including methodological considerations and the roles of embodiment, phonosemantics and oppositional relationships in an attempt to explain the semantic clusters and relations found. Additionally, this chapter

features suggestions of possible linguistic primitives. Lastly, chapter 6 provides concluding remarks and suggestions for further research.

2. Background

This chapter introduces potential universal aspects of language with a lexical focus. Relating to this the phenomenon of phonosemantics is presented as an universal factor of language and its close connection to fundamental vocabulary.

2.1 Lexical Universals

According to Saaed (2003:71), it may be expected that color terms and how they are divided should vary since cultural systems such as kinship terms and governance, as well as names for plants and animals are very different between languages. However, it seems like there are some universal features included in how color term systems constructed. Berlin & Kay (1969) investigated the variation of describing basic color terms in different languages. Their criteria for a "basic" color term to be included were that it would have to be monoleximic (not being a hyponym of another color term), having wide applicability, and that it would not be a semantic extension. They found that the number color terms varied greatly from two up to 11. Saeed writes that this might seem to support the notion of linguistic relativity, i.e. the shaping of our cognition and perception based on the language we speak cf. Slobin (1997, 2000) etc. However, within the range of each color, there is a basic focal color which was identified independent of the language spoken. It was also found that the structure of the color term systems was hierarchally built; certain colors were only lexemic as long as a specific, different color already was lexemic, Figure 2.

WHITE		→ GREEN			→ PURPLE
	→ RED		→ BLUE	→ BROWN	→ PINK
BLACK		→ YELLOW			→ ORANGE
					→ GREY

Figure 1: The color hierarchy based on Berlin & Kay (1969); the color to the left is more basic than the color to the right.

All languages had at least terms for WHITE and BLACK, followed by RED, either GREEN or YELLOW could be the forth color in a system, and the order of the very last color terms PURPLE, PINK, ORANGE and GREY varied, yielding eight kinds of basic color systems in total. Furthermore, in Heider (1971; 1972a; 1972b) it was found that when speakers of the New Guinean language Dani were compared with speakers of American English in memory tasks concerning colors, the same kind of mistakes was made. This is particularly interesting

since Dani only has two color terms; *mili* for cold, dark colors and *mola* for warm, light colors, while English has 11 terms. The conclusion which can be drawn from this is that our conception of at least the two most basic color terms seem to be affected by physiological rather than linguistic constraints, which both supports the notion of having a term for WHITE and BLACK as universal, as well as being able to distinguish the basic focal colors. These facts also open the possibilities for universal concepts and notions, and for this paper's purpose; lexical universals.

2.1.1 Swadesh Lists

In lexicostatistics and glottochronology i.e. the study of the chronological relationship of language, the so-called *Swadesh-lists* are frequently used, named after Morris Swadesh and featured in Swadesh (1971: 271-284). The principal idea of glottochronology is to count cognates among vocabularies of related languages and assuming that the more the forms diverge from each other, the longer the time span dialects have been separated. If there was no documentation of the Roman Empire and the consequences of its existence, it would still be possible to draw some conclusions of its spread throughout the Mediterranean region due to French, Spanish, Romanian, Portuguese etc. all having similar vocabularies and inflectional patterns, indicating common origin. This enables theories to be made for reconstructions of non-documented (sometimes prehistoric) languages such as Proto-Indo-European.

The words used for these types of studies were to be "basic" vocabulary, relevant for all languages of the world, excluding community-specific concepts, which according to Swadesh are more resistant to change and loss than other words. The basic word-list is then compared with divergence over time. By looking at well-document historical languages and its decedents, Swadesh established an index of for how long words roots of the list would be retained and hence relationships and age can be estimated. What is of particular interest for this study is the list of "basic vocabulary" and what criteria is used for constructing it. So called cultural vocabulary is not suitable since human work, art, customs etc. are tied to particular communities. The same applies to names of specific animals and plants since some are cultivated by man and many are only found in only few regions of the world. Swadesh continues in excluding words which are known to often be affected by phonosemantics due to the tendency of phonological forms to persist for a longer periods of time than arbitrary words do. What remains is "basic" or at least something very close to universal concepts, including very simple things, qualities and activities; such as pronouns, a few quantitative concepts,

parts and simple activates of the body, movements and general qualities such as size and color. The elements in the list has varied in number, though the more commonly used 100-item list featured in Swadesh (1971) is presented below in Table 1.

Table 1: The items of the 100-item Swadesh list.

I	root	breasts	rain (noun)
you (sg.)	bark (of tree)	heart	stone
we (inclusive)	skin (noun)	liver	sand
this	flesh	drink (verb)	earth
that	blood	eat	cloud
who?	bone	bite (verb)	smoke (noun)
what?	grease	see	fire
not	egg	hear	ash(es)
all	horn	know	burn (verb intr.)
many	tail	sleep	path
one	feather	die	mountain
two	hair	kill	red
big	head	swim (verb)	green
long	ear	fly (verb)	yellow
small	eye	walk (verb)	white
woman	nose	come	black
man	mouth	lie (recline)	night
person	tooth	sit	hot
fish (noun)	tongue	stand (verb)	cold
bird	claw	give	full
dog	foot	say (verb)	new
louse	knee	sun	good
tree	hand	moon	round
seed (noun)	belly	star	dry (adjective)
leaf	neck	water (noun)	name

Several problems with this methodology have been brought up by Swadesh himself, e.g. lack of data for some languages, while other languages are well-documented, which could blur true relationships, as could borrowing of these words between languages (cf. Haspelmath & Tadmor 2009). Also identifying cognates requires deep knowledge of phonology and structural characteristics of the languages in order to combat errors which could grow exponentially when working diachronically. 100 words might be too few in order to receive reliable statistics and to avoid effects of chance. However, according to Swadesh (1971) it would be more hazardous to include more elements simply for the sake of statistics, since the presumed universal aspect of these words is paramount. There might be difficulties in finding equivalents of the elements between languages, as mentioned by Saeed (2003:41), and furthermore non-cultural words might not even exist. Variations of connotations of terms can vary greatly, e.g. *sun*, *moon* and *fire* could be connected to deities, and *heart*, *eye* and *hand* are not always only anatomical parts, but can be connected to concepts such as *soul*,

understanding and *strength*. Despite these potential problems, Swadesh-lists are frequently used, probably due to the fact that they are limited to basic vocabulary which "does not feel any pressure to change or resist change" (Swadesh 1971:283).

2.1.1.1 Refined Lists

Based on a paper by Brown et al. (2008) describing an automated comparison procedure of word lists, Holman et al. (2008) suggest a reduced 40-item Swadesh list instead of the more conventional 100-item version, grounded in the stability of the lexical items retained over time, which according to Holman et al. (2008) yields equally good classificatory results as the larger list. Using the database of Brown et al. (2008) consisting of the 100-item lists for 245 languages, the automated classification method was tested on non-controversial language families yielding results in agreement with expert views on family sub-groupings. Over 900 languages were then examined, giving stable results for 40 out of the 100-item Swadesh list, including; BLOOD, BONE, BREAST, COME, DIE, DOG, DRINK, EAR, EYE, FIRE, FISH, FULL, HAND, HEAR, HORN, I, KNEE, LEAF, LIVER, LOUSE, MOUNTAIN, NAME, NEW, NIGHT, NOSE, ONE, PATH, PERSON, SEE, SKIN, STAR, STONE, SUN, TONGUE, TOOTH, TREE, TWO, WATER, WE, YOU (sg.). The 100-item list had an average stability of 23.4%, while the 40-item list had 30.5%, with a maximum stability of 42.8%, suggesting that the smaller list is more suited for language comparison. Furthermore, there was also indications of meanings of the 40-item list being more resistant to borrowing.

Evidence for super-families, i.e. linking established language families together distantly in the past, has been criticized due to semantic and phonetic erosion over time, yielding a maximum range of lexical stability of 5,000 to 9,000 years. According to Pagel et al. (2013), this can be overcome by finding words with sound-meaning correspondence which are conserved for such a long time that traces can still be found between separated language families. They write that words in general have around 50 % chance of being replace by another root every 2,000-4,000 years. However, some word are more conservative in nature, e.g. numerals, pronouns, and certain adverbs are replaced more slowly, around 50 % of change every 10,000, 20,000 years or more. Based on the Indo-European, Uralic, Sino-Tibetan, Niger-Congo, Altaic, Austronesian families, as well as Basque and Tok Pisin, the frequency of which a word is used in every day speech corresponds to how rapidly words evolve, i.e. frequently used words evolve slower than seldom used words (Pagel et al. 2007). Using a core set of stable everyday speech possible cognates shared between the Altaic, Chukchi-

Kamchatkan, Dravidian, Inuit-Yupik, Indo-European, Kartvelian, and Uralic language families were examined, which traditionally are regarded as unrelated. The results showed that words used more than once per 1,000 in everyday speech had such a slow lexical replacement rate that there was a high chance of finding them as cognates in more than two of the language families. Cognates shared by four language families or more were considered reliable, resulting in 23 words, mainly numerals, pronouns, and special adverbs, Table 2. It was also found that frequently used words were at least seven times more likely to be judged as cognates that infrequent words.

Table 2: The 23 proposed cognates of the Altaic, Chukchi-Kamchatkan, Dravidian, Inuit-Yupik, Indo-European, Kartvelian, and Uralic language families, occurring in more than three families.

Number of cognates shared among language families			
7	6	5	4
YOU (sg.)	I	NOT, THAT, WE, TO GIVE, WHO	THIS, WHAT, MAN, YOU (pl.), OLD, MOTHER, TO HEAR, HAND, FIRE, TO PULL, BLACK, TO FLOW, BARK (noun), ASHES, TO SPIT, WORM

YOU (sg.), I, NOT, THAT, WE, TO GIVE and WHO were suggested to probably being ancient based on controlling the likelihood of chance sound associations. Possible alternative explanations for the cognates such as borrowing to lack of exchange between the language families etc. were disregarded. Pagel et al. also mention the possibility for some types of words being more likely to appear as cognates than others. For this argument, they propose closed-class words of simple phonology, i.e. all of the word being shared by more than four families, whose short length could contribute to creating similar proto-words. Though in sum they found a sizable list of possibly related words from a common ancestor around 15,000 years ago.

2.1.2 Natural Semantic Metalanguage

Anna Wierzbicka and Cliff Goddard have developed numerous studies of basic concepts. Goddard & Wierzbicka (2002) write that if it is possible to perform a semantic analysis using reductive paraphrasing, without circularity, what is left is the semantic core of language, which is judged by Geeraerts (2010:127-134) to be the most serious attempt to find an inventory of universal primitive concepts. This core must have a language-like structure, contain a lexicon of indefinable expressions (called *Semantic Primes*) from which all other expressions can be built, and principles governing how they can be combined (grammar). This constitutes a miniature-language with the same expressive power as a natural language, which Goddard and Wierzbicka call *Natural Semantic Metalanguage*.

In order to find the semantic core of language, a method of trial and error has been employed by Wierzbicka since the 1970s. The desired goal is to find the smallest and most versatile set of concepts, which currently includes 60 items. These are however not considered finite; Goddard stresses the importance of the framework as being work in progress. The first criteria for the semantic primes is that they are indefinable, i.e. that they cannot be paraphrased in any simpler terms. This in turn means that definitions of concepts should be able to be written in natural, non-technical language instead of a formalized representational language, i.e. the semantic primes ought to be able to paraphrase every other word. Table 3 below shows the semantic primes in English, although the English forms should only be viewed as vehicles for explaining the potentially universal concepts. For example in certain languages some concepts have more than one form, e.g. English MUCH and MANY, which in the natural semantic metalanguage theory are referred to as an *allolexy*, but do denote the same concept.

Table 3: The 60 proposed semantic primes in their English forms.

Category	Primes
substantives	I, YOU, SOMEONE, PEOPLE, SOMETHING/THING, BODY
determiners	THIS, THE SAME, OTHER
quantifiers	ONE, TWO, SOME, ALL, MUCH/MANY
evaluators	GOOD, BAD
descriptors	BIG, SMALL
mental predicates	THINK, KNOW, WANT, FEEL, SEE, HEAR
speech	SAY, WORDS, TRUE
actions, events and movement	DO, HAPPEN, MOVE
existence and possession	THERE IS, HAVE
life and death	LIVE, DIE
time	WHEN/TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME
space	WHERE/PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE
logical concepts	NOT, MAYBE, CAN, BECAUSE, IF
intensifier, augmentor	VERY, MORE
taxonomy, partonomy	KIND OF, PART OF
similarity	LIKE

All primes are very simple and have intuitively intelligible meanings which are grounded in ordinary linguistic experience. The oldest members of the 60 primes list dates back to Wierzbicka (1972) and includes I, YOU, SOMEONE, SOMETHING, PART, THIS, SAY, HAPPEN, WANT, FEEL which are also the most intensively investigated elements. Other elements have either been added or tweaked, e.g. in the original version the element NOT was represented by *don't want* (*diswant*) and IF by *imagine*. The rest of primes have been added in two waves.

After presenting the viability of the primes and the metalanguage for English, Goddard writes that if semantic primes were established for each and every language, the lists would turn out very differently due to polysemy and differentiations in grouping concepts. However what is interesting is which elements would be present in all languages. Hence the second criterion of the semantic primes is that they ought to have analogues in all languages, referred to as the *Strong Lexicalization Hypothesis*. If the primes are a core part of language, the analysis made for English should hold for all languages of the world. Goddard lists 26 genetically and areally diverse languages, in which in-depth analyses have been made using the framework; English, Polish, Italian, French, Ewe, Amharic, Mandarin, Cantonese, Lao, Thai, Malay, Acehnese, Japanese, Kalam, Mangaaba Mbula, Kayardild, Longgu, Bunuba, Arrernte, Yankunytjatjara, Hawaiian Creole English, Samoan, Spanish, Tsimshian, Misumalpan and Maori. The results seem to indicate that some concepts are very common part of the lexicon of (maybe all) languages and perhaps a fundamental part of the language faculty. Though the primes are "lexical" in a very broad sense, they could be realized as a phrasemes in one language, and bound morphemes in another. However, according to Goddard, as long as they convey the correct meaning the results of the framework ought to be reliable.

There is of course plenty of criticism for this rather bold theory, of which some is presented in Geeraerts (2010:133-134). According to Wierzbicka and Goddard, the set of primitives change over time through trial and error which would make it quite empirical in nature, though Geeraerts points out that there is no indication of what "error" actually means. He further points at that Wierzbicka & Goddard's method for establishing primes is practical, though only useful if a sufficient number of languages is used. There is also criticism from individual languages lacking the proposed primes, such as Bohnemeyer (2003) who argues, e.g. BEFORE and AFTER not being lexical in Yukatec Maya.

2.1.3 The Oppositional Relation

Lévi-Strauss (1955, 1984) summarized by McTurk (2005:314) writes that a myth is an attempt of a society to try to reduce and resolve contradictions and paradoxes of the world as it is perceived on rational grounds. Hence myths are built up by oppositional ideas, i.e. life and death, heaven and earth, human and animal, order and chaos and so forth. Several oppositions can be presented within the same myth, creating more oppositions on deeper level, though the actual message of the myth is centered around one fundamental opposition. It is possible that oppositional thinking is fundamental to cognition, though it is seldom

completely clear-cut. Firstly, a single concept can be associated with various notions (often dependent on culture), e.g. the color red can be connected to love and passion, but also to danger. White is the color of mourning in some cultures, while its opposite, black, is used for this purpose in other cultures. Secondly, oppositional pairs can be created depending on what is contrasted. If a penguin is grouped with a pigeon, an eagle and a hawk, it would probably be judged the odd one out due to not possessing the ability to fly. However, if the hawk is substituted for a cat, most would group the penguin together with the pigeon and the eagle on the ground that the cat is not a bird at all, Figure 3.

penguin	pigeon
eagle	hawk

penguin	pigeon
eagle	cat

Figure 2: Examples of shared properties depending on context between different animals, red indicating the odd one out.

Willners (2001:59-72) writes that so called "direct" antonyms are rather easy to find in Mandarin Chinese compared to many other languages, including Indo-European, due to a specific rule of compounding which only applies to antonyms pairs. By combining two antonymous adjectives a noun is created, which is a quality referring to both of the bipolar extremes, e.g. *hǎo-huài* (lit. 'good-bad') meaning 'quality', and *dà-xiǎo* (lit. 'large-small') meaning 'size' (Li & Thomson 1981:80). Another illustrative example are the young men of the aboriginal Walbiri people of Central Australia who are taught the secret language Tjiliwiri after entering adulthood, explained by Hale (1971). This language is a kind of mirror image of Walbiri, i.e. all nouns, verbs and pronouns of ordinary Walbiri are substituted by their antonyms. Hence 'you are tall' is expressed by 'I am short', and 'I am sitting on the ground' by 'you are standing in the sky'. According to Willners this shows that some people are explicitly aware of the semantic relation of antonymy.

Some studies concerning speech errors seem to indicate that closely related words such as antonyms are frequently uttered when the desired word fail to come through, so called slips of tongue, e.g. Willners mentions Söderpalm (1979) and Linell (1982). Other ways of elicitating closely related words include psycholinguistic experiments such as word-association test. Deese (1965) used a type of stimulus-response test, where words function as stimuli and the subjects were told to say the first word which came to mind upon hearing the stimuli. Words with similar meaning evoked the same responses, showing a strong relation between opposites in an antonymous pair. Justeson & Katz (1991) used an English corpus in order to show that

antonymous concepts occur in the same sentence more frequently than chance. By studying 35 antonymous pairs previously identified by Deese (1965), Table 4 below, they concluded that antonym co-occurrence is 8.6 times more frequent than expected and concluded that "co-occurrence takes place via substitution, substitution yields antonyms alignment, and alignment leads to association". By reexamining the same 35 word pairs Willners (2001) found that for most pairs, co-occurrence was significant; 25 at a significance level of 0.05, 19 at 0.01 and 14 at 10^{-4}, and by using a different corpus all 35 pairs co-occurred more frequently than expected, showing the cognitive relation between the co-occurring antonyms. *Old* was however found to be a true antonym of both *new* and *young*, *good* was an antonym of both *bad* and *evil*, both *large-small*, and *big-little* described the field SIZE.

Table 4: The 35 antonymous pairs identified by Deese (1965).

big-little	fast-slow	left-right	sour-sweet	far-near
cold-hot	heavy-light	new-old	strong-weak	happy-sad
dark-light	thick-thin	poor-rich	deep-shallow	rough-smooth
high-low	narrow-wide	right-wrong	easy-hard	short-tall
large-small	back-front	active-passive	alive-dead	bottom-top
long-short	bad-good	dry-wet	clean-dirty	inside-outside
old-young	black-white	hard-soft	empty-full	pretty-ugly

Willners proceeded by conducting the same study using Swedish, though due to translation and usage difficulties of some of the antonyms investigated by Deese and Justeson & Katz, e.g. 'far' could be either translated *fjärran* or *långt borta* (lit. 'long away'). Another but similar list of antonymous adjectives was defined by Lundbladh (1988). All words co-occurred more than expected when tested by using yet another corpus. Excluding the field of SIZE the antonyms co-occurred 3.17 times more than expected. In a later study made by Paradis, Willners & Jones (2009) canonical antonyms were investigated, i.e. a small set of words with special lexico-semantic attraction which are entrenched in memory and perceived as strongly coupled pairings by speakers. 8 antonym pairs were considered canonical; *slow-fast*, *light-dark*, *weak-strong*, *small-large*, *narrow-wide*, *bad-good*, *thin-thick* and *ugly-beautiful*. It was found that antonyms took significantly longer to process than synonyms, while there was no significant difference between the response times of synonyms and unrelated word pairs. However, canonical antonyms were significantly faster to process than the non-canonical antonyms, indicating a close relationship between the canonical antonym pairs. It was also found that by using an elicitation experiment, for 7 canonical antonyms, only one and the same antonym was suggested by the subjects; 'bad' ('good'), 'beautiful' ('ugly'), 'clean' ('dirty'), 'heavy' ('light'), 'hot' ('cold'), 'poor' ('rich') and 'weak' ('strong'). It was concluded that for all

case studies, direct antonyms co-occur notably more often than chance would predict, between 3.17 to 7.0 times more, and even though these were constricted to English and Swedish, two very closely related languages, the results tells us something about how we cognitively treat opposition in words and concepts.

2.2 Phonosemantics

Starting over 2,000 years ago in both ancient Greece and ancient China, the existence and the behavior of phonosemantics, as well as its role in language evolution have been debated. Plato wrote in his famous dialogue *Cratylus* that the sound [l] would be better suited for words representing liquid meanings. This was later continued by Herder's (1772) "Essay on the origin of speech", arguing for a phonosemantic origin of language. The dogma of total arbitrariness which has been prevalent in much of the 20th century originates from Saussure (1916), who considered language to have no need for non-arbitrary connections between the *signifier*, the form of words, and the *signified*, the concepts in question, since concepts do not precede language. However, contemporary with Saussure and at the opposite side of the argument, we find Jespersen (1922) who wrote that onomatopoeic words seem to resist sound change, exemplified by the vowel change from [u] to [ʌ] in *cut*, which has not occurred in the imitative word *cuckoo*.

In many of the world's languages, complete word classes exist which have a direct relationship between sounds and meaning, referred to as *mimetics* in Japanese and Korean, *expressives* in Austro-Asiatic languages and *ideophones* in African languages, the latter perhaps being the more commonly used term of the three. Diffloth (1994) describes this phenomenon to be an attempt by speakers to convey various sensations as directly as possible through language, such as color, sound or movements etc. In many cases the meaning of these types of words is understood by most speakers, though due to the filter-like function of language, when it comes to the use of the vocal apparatus to imitate various sounds and events of the world, many ideophones can become very language specific; meaning that a motivated relationship between sound and meaning has become at least partially arbitrary. Similarly *phonesthemes*, e.g. initial "*gl-*" in several English (as well as other Germanic) words such as ***gl****immer* and ***gl****itter*, are referred to as *conventional sound symbolism* by Hinton, Nichols & Ohala (1994). These analogical associations of phonemes or phoneme-clusters can be used to invent new words, whose meanings are understood by speakers of these languages without previously hearing them uttered. Based on these cases, one could argue that phonosemantics

is a case of analogy or language specific conventions. However, since Jespersen several experimental studies on phonosemantics have been conducted showing supporting results for cross-linguistic patterns, summarized in Abelin (1999).

2.2.1 Phonosemantic Experimental Results

One of the most famous phonosemantic experiments conducted is the one by Sapir (1929) in which he asked 500 subjects to choose between the words *mil* and *mal* to be the more fitting name for a small table and a large table. 80 % agreed that *mal* was the more fitting term for describing the large table. Following Sapir's experiments, Newman (1933) tested both vowels and consonants for associations with small-large and bright-dark. He found that small/bright-large/dark agree with front-back vowels and with labial-dental consonants. In a study conducted by Bentley & Varon (1933), it was indicated that [a]-sounds were felt to be larger, rounder and softer than [i]-sounds. Usnadze (1924) and Köhler (1930), later continued by Ramachandran & Hubbard (2001), all investigated the relationship between sounds and shapes by asking subjects to match the two nonsense words *maluma* and *takete* with drawings, of which 95 % matched *maluma* to the rounded figure and *takete* to the angular figure. This also further developed by Ahlner & Zlatev (2010), who found more precise sound-meaning associations; a pointy (also described as "hard" by participants) shape was grouped with [i] and the voiceless obstruents [p, t, k, tʃ], while a roundish (also described as "soft" by participants) shape was described as being best associated with [u] and the voiced sonorants [m, l, n, ŋ]. This was true for for 90 % of subjects when concerning vowels and 80 % concerning the consonants. In addition, when "round" vowels were grouped with "pointy" consonants and vice versa, the results showed that the sound sequences were considered ambiguous. Wisseman (1954) investigated the creation of onomatopoeic words from noises by presenting noises to subjects, who were then asked to invent names for them. It was found that [i] imitates high-pitched noises; [u] imitates low-pitched noises; words beginning with voiceless plosives ([p], [t], [k] in this case) imitate noises with abrupt beginnings, while gradually starting noises were described by words beginning with [s] or [ts]. Chastaing (1958, 1965, 1966) investigated French and found that [i] was connected to acuteness, smallness, lightness, rapidity, and closeness; stops are hard; continuants soft; [r] is rough, strong, hard; [l] is smooth, weak, light-weighted. Fonagy (1963) compared [i] and [u] in Hungarian by studying both children and adults. [i] was perceived as being quicker, smaller, prettier, friendlier, harder than [u], while [u] was perceived as thicker, hollower, darker, sadder, blunter, bitter and stronger. It was also found that [r] was perceived as wild, pugnacious,

manly, rolling and hard. Sereno (1994) wandered away from sense-related associations of sounds and focused on lexical organization instead; the most common in English verbs and nouns. Through a reaction time experiment the results showed that verbs with front vowels (lexically most frequent) were recognized faster than verbs with back vowels. And reversely, nouns with back vowels (lexically most frequent) were recognized more quickly than nouns with front vowels, showing a possible correlation between sound usage and word class.

Cross-linguistic studies include Osgood et al. (1957) who investigated the dimensions of 'value', 'strength' and 'potency' by rating nonsense syllables on the semantic differential. The results of both English and Japanese speakers were almost identical and generally showed that front consonants were found to be more pleasant than back consonants and high frequency sounds are associated with smallness and impotence. Furthermore, studies concerning prosody and emotions have been conducted by Wilde (1958), Müller (1960) and Abelin & Allwood (1984). Ultan (1978) found phonosemantic associations connected to the distance encoded in demonstrative systems, by looking at 136 languages; having more closed, fronted, unrounded vowels in proximal words, as well as in diminutive markings. Woodworth (1991) showed evidence for a similar relationship between vowel quality and demonstrative pronouns (including locative adverbials). Out of 26 languages, 13 had linguistic forms in which vowels with higher F2 frequency were used for the proximal form than for the distal form, with only two languages showing the opposite relation. This was later reinforced by Johansson & Zlatev (2013) who investigated several possible motivations of phonosemantics for deictic demonstrative pronouns, finding that the frequency of vowels was the most important factor. Woodworth's investigation was expanded by Traumüller (1994) who also found support for the occurence of vowels with higher F2 frequency in proximal demonstratives compared to their distal counterparts in 32 out of 37 languages. Furthermore, Traumüller investigated first and second personal pronouns for sound-meaning associations. For hypothesis 1 it was found that in 11 out of 14 cases first person was expressed by voiced nasals, while second person was expressed by a voiceless stop. For hypothesis 2, it was found that in 9 out of 15 cases, first person was expressed by sounds with an absence of dental articulation and/or lip protrusion e.g. [x], while second person was expressed by sounds with a presence of dental articulation and/or lip protrusion e.g. [t] or [w].

Swadesh (1971:157-212) writes that different rhythms of syllables could symbolize particular types of movement, while largeness could be shown by broad vowels, smallness by narrow

and roundness by round vowels. Sibilants could be associated with rustling and rushing as well as rubbing and slipping, while laterals could be connected with lightness, glancing contact and flexibility, and rhotics with heaviness, crashing impact, hardness and harshness. Voiceless stops could convey hard objects and impacts, while voiced stops could convey soft impact, also nasals reflect relaxation, contentment and resonance. Further, Swadesh lists aspiration to convey energy, resistance and force, voicing to convey weakness and non-opposition also yielding friendliness and affection, clicks and implosives to convey movements toward speakers, fricatives to convey rubbing, and glottalization to convey quickness and abruptness. These sound types can then be combined into clusters yielding combined meanings e.g. laryngeal constriction combined with nasals indicates displeasure or frustration, cf. phonesthemes.

2.2.2 Phonosemantics for Cross-Linguistic Comprehension

Defending the posiion of the possible universality of phonosemantics is the fact that certain novel words can be understood based on their sound makeup. Tsuru & Fries (1933) used lists of opposite word pairs in two different languages such as English *large/small* and German *gross/klein*, presented orally. The subjects spoke only one of these languages and were asked to identify the corresponding meanings of the words of the unknown language, yielding results exceeding chance.

LaPolla (1994) conducted a similar investigation. He claims that in many East Asian languages a change to a high tone or rising tone is used to mark diminutive or familiar, though in Mandarin Chinese hypocoristic forms of nouns are marked by affixation of retroflex suffix and no change of original tone. The second semantic use of a high tone in Cantonese is to increase intensity of reduplicated adjectives. In order to mark extreme intensification in Cantonese, Thai and the Hainan form of the Southern Min dialect of Chinese, the first syllable of a word is modified to create a high-low tonal pattern. Furthermore a low-high pattern marks familiarity.

LaPolla presents a series of conducted experiments concerning phonosemantics, the first one being English speakers matching 40 Chinese antonymic word pairs to English translations correctly, yielding the results; 190 correct answers and 169 incorrect answers ($p < 0.30$). This was followed by 10 English speakers performed the same test though with the different tones of the Chinese words in reversed order to investigate whether the tones were instrumental for

making judgments, yielding 187 right answers and 150 wrong ($p < 0.05$). According to LaPolla, the similar results indicate that tones could be an important criterion for making judgments, i.e. that the change in tone could make it easier to give the correct answer. A high level tone was favored for "small"-category words, hence *wide* and *course* (high tones) received many incorrect answers. A falling tone was favored for "large"-category words, except for *soft* for which LaPolla suggests that the "large"-category could consist of several subgroups.

Then, an experiment testing native Mandarin speakers' sensitivity to sound-to-meaning was presented by matching 48 oppositional meanings with 50 nonsense syllables of possible Mandarin syllables and assign tones to the words in a way they found fitting. Grave (low frequency energy) segments were matched with "large"-category words in 151/210 ($p < 0.001$) cases but only in 103 cases for "small"-category words. Nasal segments were also measured showing that for "large"-category words [m] occurred in 17 cases and [n] in 6 cases, while for "small"-category words, [m] was less frequent, occurring 9 times, while [n] occurred 20 times ($p < 0.01$for both). Also, high tones were more frequent for the "small"-category words and falling tones for "large"-category words, which according to LaPolla correlates with the findings of Yue-Hashimoto (1980) and Li & Thomson (1977) in which children learned high and falling tones, before other tones when acquiring Mandarin as their mother tongue.

The native Mandarin speakers were also studied by having 25 Cantonese words or phrases which could be either diminutive/familiar or intensified by changing to a high or high-rising tone, and asking the Mandarin speakers to match the words or phrases with default and modified tones. No significant results were found for the words concerning familiarity; 21/40 correct answers. However for the hypocoristic formations 87/150 correct answers were given ($p = 0.05$), and for intensification 15/20 correct answers ($p < 0.05$). Lastly, very strong results was found for extreme intensification; 34/40 correct answers ($p < 0.001$). LaPolla concludes that generally the results show a tendency to associate acute (high frequency energy) segments with "small"-category words and grave segments with "large"-category words.

2.2.3 Phonosemantic Sounds and Semantic Domains

Generally, the semantic domains featured above include notions of acoustic sound, form, luminosity, wetness, weight, thickness, size, distance, events and movements, evaluative

attitudes, deixis, surface, softness, sharpness and so forth. These yield the more general domains of hearing, vision, touch, movement, form, mind, attitude (including evaluative), size, number, deixis and class. However, Abelin (1999) notes that the semantic domains occurring include the senses of 'hearing', 'vision' and 'touch', while "taste' and 'smell' is missing. Also, 'movement' is often co-occurring with the perception of 'sound' and 'form' is connected with 'vision' as well as with 'touch'. Abelin continues with suggesting that while 'mind' (attitudes and emotions) might be indexically connected to expressions for dislike, pejorative expressions and so forth, 'size' can be seen as iconically related to sound. Furthermore, the domains are usually contrasting ends of scales, i.e. semantic oppositions such as deictic expressions. Phonetically, vowel features include the dimensions of opened/closed, front/back and rounded/unrounded as well as specific vowels. Consonants seem to be treated in larger groups such as obstruents and sonorants, but also in smaller groups connected to manner of articulation i.e. fricatives, stops, nasals etc. Furthermore the semantic domains investigated usually have binary poles, e.g. *large-small*[1] (large size or extent and small size or extent) and *here-there* (large distance and small distance), which fits the vowel dimensions, but also in many cases different groups of consonants.

The connection between a sound and a meaning is by no means exclusive, i.e. a sound can denote several different meanings, often depending on contrast; an [e] can be perceived as smaller than [a] while it can be perceived as larger than [i], cf. the results concerning deictic demonstratives in Johansson & Zlatev (2013) and Traumüller (1994). A sound's different associations do not have to be within a single semantic domain, much like the English word *light*, which can mean both pale in color and low in weight; a sound can act as one of the poles of many basic oppositional pairs. Swadesh (1971:157-181) stresses that what we might perceive as simple animals cries, is often produced by humans as well, in contexts of being hurt, frightened or under other strong emotional states. The tone and quality of the sounds helps in warning others of potential danger etc. A normal speech melody indicates that the speaker is relaxed, rapid speech and sharp variations in tone mean that he or she is excited and various rhythms and emphases can convey everything from friendliness or anger. For further details of biological explanations and associations with intonation see Gussenhoven (2001). Closely related to this are exclamatives, which not necessarily follow a language's phonology,

[1] Note that English *big* and *small* show the direct reversed relationship between vowel value and meaning to what would be expected, however, for each phonosemantically motivated concepts there is probably one or two languages which show opposite, though the dominant pattern still show consistant correlations.

e.g. English (also present in many other languages) *pst*, and *shh* lack vowels, not permitted in English phonology, nasalized vowels can be found in utterances such as *aha* and *uh-uh*, and click sounds are found in e.g. *tsk-tsk*. Swadesh (1971:157-181) calls these sound-meaning associations *vocal gestures*, perhaps more often used when sounds or textures are hard to symbolize using manual gestures. And these, he argues, are both important and common when people who do not share the same language are attempting to communicate with each other, showing their describing potential. Many gestures do however vary, e.g. pointing with lips or fingers as well as gestures directly showing the desired meaning are probably more or less universal, while e.g. negation can be indicated by shaking of the head, nodding, wagging a finger and so forth, depending on culture. This division is also similar to what could be called universal phonosemantics, i.e. cross-linguistic and phonesthemes described above, usually occurring in one language or within a language family.

2.2.3.1 The Frequency Code

Ohala (1994) coined the so called "frequency code" which puts phonosemantics in a larger ethological context, based on the fact that certain meanings seem to be reflected in vocalic sounds and certain facial expressions of humans and other species' communication systems (see also section 3.3). Threatening sounds of most animals such as a dog's growl are low in frequency while a dog's whine, denoting e.g. submissiveness is high-pitched. The explanation for this correlation between sound and meaning probably has many intertwined layers and levels, but mainly, functional reasons probably account for this. In many cases it is in animals interests to appear large, since that would mean that they would be perceived to have upper hand in potential confrontations. This can be achieved visually by erecting hair or feathers, or using lower than normal vocalizations based on that larger animals have larger resonance chambers which produce lower frequency sounds. Doing the opposite, i.e. manipulating the body or vocalizations to seem smaller, can be equally functional when confronted with an unbeatable opponent.

According to this theory, a high and/or rising F0 are associated with smallness as well as with related concepts such as deference, politeness, submission, lack of confidence, questions, familiar, dependence and narrow, near, while a low and/or falling F0 is associated with largeness but also assertiveness, authority, aggression, confidence, threat, dominance, statements and large distance. Along with the frequency of the F0, the frequencies of the other formants, the most important being F1, F2 and F3, are affected by the three vowel dimensions opened/closed, front/back and rounded/unrounded. Furthermore, general universal tendencies

of sound patterns used for certain meanings are listed by Hinton, Nichols & Ohala (1994); stops with abrupt sounds and acts, continuants with continuing sounds and acts, fricatives with quick audible motions through air and nasals with ringing and reverberating sounds.

2.2.3.2 Sonority

Cross-linguistically, there are restrictions on how onsets and codas are built up by segments. If a cluster of two consonants is allowed in the coda or onset position, the same position has to permit a single consonant in the position as well. Similarly voiced versions of plosives in a language generally imply the voiceless versions of the same sounds (Abrahamsson 2004). Also, many languages only allow sonorants in the coda position while other also allow obstruents, the reversed way however does not occur, meaning that obstruent in coda position implies at least one sonorant possible for that position. Abrahamsson also mentions studies indicating that certain types of sounds in the coda and the onset are more or less difficult to learn, though there is no real consensus on the area. The positions of actual types of speech sounds also follow certain patterns cross-linguistically in terms of sonority, i.e. the grade of intensity different segments produce depending on the resonance of the speaking tube, which are dependent on the degree of obstruction of the airstream. Commonly, speech sounds are for this purpose divided into vowels (most sonorous), semi-vowels, liquids, nasals, fricatives and plosives (least sonorous) yielding the so called sonority hierarchy (see also section 3.3). Hogg & McCully (1987) propose a more fine-grained variant with sonority values for each sound group, Table 5. The main driving force is still the degree of obstruction of the airstream, thus high vowels are less sonorous than low vowels and nasals are more sonorous than corresponding oral stops etc. Voicing also increases sonority, meaning that voiced fricatives are more sonorous than voiceless fricatives, but less sonorous than nasals. Hence an optimal syllable ought to have a nucleus consisting of a vowel, with gradually less sonorous segments towards the edges of the onset and the coda.

Table 5: The sonority values of different types of sounds based on Hogg & McCully (1987:33).

Sounds	Examples	Sonority value
low vowels	[a, ɑ]	10
mid vowels	[e, o]	9
high vowels	[i, u]	8
flaps	[r]	7
laterals	[l]	6
nasals	[n, m, ŋ]	5
voiced fricatives	[v, ð, z]	4
voiceless fricatives	[f, θ, s]	3
voiced plosives	[b, d, g]	2
voiceless plosives	[p, t, k]	1

However, Abrahamsson writes that the hierarchy is not a cross-linguistic law, but rather significant phonotactical tendency. In many languages, including English /s/ violates the sonority hierarchy in words such as apparent in /st/-, /sk/-, /sp/-clusters e.g. *stick* and *scout*. Furthermore, many African languages have phonologies allowing nasals to occupy the very edge of syllables despite being highly more sonorous than plosives and fricatives. Abrahamsson (2001, 2003a) also found that in second language acquisition, Chinese learners of Swedish found it more difficult to learn words which did not follow the sonority hierarchy, though with many exceptions. The grade of obstruction could possibly be correlated with some meanings, e.g. the roughness-smoothness dimension which could correlate with low and high sonority.

2.2.3.3 Articulation and Acoustics

By conducting an investigation of phonosemantics for 'small' or 'large' through a rating experiment that included speakers of Chinese, English, Japanese, and Korean, Shinohara & Kawahara (2012) claim that the height of vowels, the backness of vowels, as well as voicing in obstruents, contribute in mapping sounds to meanings (Johansson & Zlatev 2013 present a similar discussion about the acoustic dimensions of phonosemantics). Shinohara & Kawahara continue by looking at the potential articulatory and acoustic explanation for these associations. Concerning the vowels, an articulatory explanation would be that lower vowels are produced by having a larger sub-oral cavity in front of the tongue which could result in a connection to largeness, while an acoustic explanation would be Ohala's frequency code hypothesis, i.e. size of resonator or resonating cavity implying size of the producer of the sound. Regarding the role of voicing in consonants, articulatorily speaking, they suggest that when producing voiced consonants several articulatory maneuvers are used to expand the oral cavity, such as larynx lowering, velum raising, and cheek expansion, which could yield a sensation of largeness. However when it comes to acoustics, they suggest that since vowels have lower F0 next to voiced obstruents than to voiceless obstruents, largeness could be implied, and adding to this is the fact that voiced obstruents have F0, while voiceless obstruents do not. A problem of the acoustic view is that low vowels have higher F1 while suggesting largeness, and furthermore deaf children are sensitive to articulatory gestures in regard to phonosemantics. However, an articulatory explanation is not enough to explain why high tones can be associated with smallness and hence Shinohara & Kawahara leave a definite answer for which of the two explanations is more contributing aside. Regardless of whether

articulation or acoustics is the more correct one for this particular case, both are probably important in understanding phonosemantics. It is apparent from the studies presented above that besides the role of acoustics in sound-meaning associations, the manner of articulation could also evoke thoughts of e.g. texture and shape. Hence it appears that certain meanings ought to be more reliant on articulation while some are more reliant on acoustics in meaning to sound mappings.

2.2.4 Phonosemantics in Basic Vocabulary

Wichmann, Holman & Brown (2010) investigated the possible role of phonosemantics in basic vocabulary by looking at approximately 3,000 of the world's living languages. 18 out of 121 language families and 52 out of 123 isolates and unclassified languages were represented. Pidgins, creoles, and mixed languages were excluded due to their ambiguity when it comes to classification. The investigated concepts were the 40-item subset of the Swadesh list featured in Brown et al. (2008). Hence the concepts were not selected for their susceptibility to phonosemantics, but for their phonological stability across time. The concepts were then transcribed according to a simplified phonological system (the ASJP-transcription system, consisting of seven vowels and 34 consonants), which was supposed to capture the most common points and manners of articulation, also following Brown et al. (2008).

Three main associations were investigated. Sound-Sound Associations; whether the distributions of different sounds over different positions in the words corresponded in the list of 40 concepts used could indicate sound symbolic effects. It was found that the vowels and consonants were not clearly segregated, meaning that there was no real tendency for vowels to correlate with each other, or for consonants to correlate with each other, any more than for vowels to correlate with consonants. Meaning-Meaning Associations; whether there was differences in meanings based on the pattern of sounds over different positions in their words. Most of the investigated concepts clustered in a dense mass, corresponding to the average pattern of sounds across all concepts. However, a possibly significant and thereby very interesting cluster found, where the pronouns I, YOU, and WE were joined by NAME. Lastly, Sound-Meaning Associations; whether the patterns of sounds in the investigated words would differ from the average pattern of sounds across all concepts. Wichmann et al. concluded that the kinds of words that tend to be inherited, not borrowed or innovated, do not correspond to those kinds of words that tend to show phonosemantics.

From these results they created what so-called "Ninatic" words, i.e. the average relative frequency of each sound in the words for each concept were calculated separately for each position in the word and was then used for creating four-phoneme words. The only meanings found with phonologically distinctive word shapes were BREAST, I, KNEE, YOU, NOSE, NAME. The first three segments of BREAST (*muma*) contain sounds articulated with the lips, which Wichmann et al. claim to be reflecting the suckling of a child. I (*naa*) and YOU (*nin*) both use *n*, however they differ in the vowels *a* and *i*. Wichmann et al. suggest that this deictic and phonological contrast to be associated with different gestures of the tongue; in [a] tongue rests in a neutral position, while it is moved forward in [i], almost pointing cf. the Vocal-Pointing motivations of Johansson & Zlatev (2013). WE (*nina*) combines the sounds of YOU and I, which is interesting since according to Wichmann et al. whenever languages had a contrast between inclusive and exclusive WE, the inclusive form was used in the study. Furthermore Wichmann et al claim that it is cross-linguistically common to construct the first person plural inclusive pronouns from the first and second person singular pronouns. The form of KNEE (*kokaau*) was suggested to possibly have to do with combination of the qualities hard and round which associates with *k*, *a* and round vowels such as *o* or *u*. Also, the Ninatic word for BONE has a similar form; *kaka*, and that according to Bloomfield (1895) *kVkV* structures often affects so-called "congeneric" classes of words. For NOSE (*nani*) they simply state that it contains two nasals. Also, an association was found, linking NAME together with the congeneric concepts I (*naa*), NAME (*nani*), and perhaps PERSON (*nanaa*). Wichmann et al write that these are so closely associated that it might be possible that they were homophonous in some primitive stage of language.

The phonologically prototypical "Ninatic" lexicon does not directly correspond to a reconstructed proto-language, but might be as close as one can get. Also, a problem arose with the Ninatic forms since [a] is often the most frequent even in the third and fifth positions, hence not always leaving room for more elaborate interpretations. What could be said about this linguistic system from their findings is that there would be a preference for CV syllables and disyllabic words, and that from basic phonosemantic building blocks other words would be constructed by using the same sound patterns for semantically related concepts, which could then be internally contrasted by other sound patterns. When the habit arose of combining sounds in conventional ways in order to denote meanings there would be, they argue, already a degree of arbitrariness involved, showing that conventionality does not equal arbitrariness cf. Ahlner & Zlatev (2010).

3. Method

In order to investigate what kind of sounds are more common in which concept, word pairs (in contrast to Wichmann et al. 2010) cross-linguistically common are used, and in contrast to Swadesh's aim of excluding words prone to be influenced by phonosemantics, this investigation will feature such concepts.

3.1 Language Sampling

For this paper, 75 genetically and areally spread out languages are used. According to the classification made by Ethnologue Online (http://www.ethnologue.com/), there are currently 6909 living languages in the world. This classification can be and has been criticized i.e. the distinction between dialects and languages, however, the 6909-estimate is judged reliable enough to use for this particular investigation. In order to reach the desired goal of 75 languages representing all the world's languages, the number of languages contained by the various language families was converted into percentages and then modified to yield 75 languages in total.

Having only 75 languages at my disposal results in language families containing less than 92 languages not being large enough to be represented by 1 language in the sample. The remaining language families were divided into five larger groups, following Bybee, Perkins & Pagiuca (1994), described by Veselinova (2005); language families containing fewer than 7 languages, isolates and unclassified languages; language families containing 7-20 languages; language families containing 21-44 languages, language families containing 45-67 languages and language families containing 68-92 languages. Languages chosen from these groups were geographically spread into six different portions of the Earth following Nichols (1992); North America (including Central America), South America, Europe (including Caucasus), Africa (including the Middle East, except Turkey), Asia and Pacific (including Oceania and the Indonesian region). <7 languages South America (1), Europe and Asia (1); 7-20 languages South America (1), North America (1) and Europe (1); 21-44 languages South America (1), North America (1), Europe (1), Africa (1), and Asia (1); 45-67 languages South America (3), North America (1) and Europe (1); 68-92 languages Africa (1) and Asia (1).

The Austronesian and Niger-Congo language families are very large and thus represent a quite large portion of the total 75 languages, hence the representing languages for these families are reduced (Austronesian from 13 to 12 and Niger-Congo from 16 to 12) in favor of

featuring more, smaller languages families for greater diversity. Generally it is rather difficult to acquire information from the Trans-New Guinea "family", and only 1 language was found to have enough information to be used for this study, hence 4 representing languages were lacking. The language family group containing 68-92 languages included Creoles, Tai-Kadai and Dravidian only, and due to the challenge of acquiring language data from the Dravidian languages, mainly because of transcription difficulties, the language family was dropped, thus the language family group containing 68-92 languages were represented by only 2 languages.

These adjustments from the original sample caused ten languages to be lacking. The empty spaces were filled by choosing languages from language families spoken in geographic positions which would, combined with the groups of smaller language families, yield a balanced geographic distribution, Table 6. The replacement languages were 1 South American, 3 North American (one being Arctic), 2 European, 3 Asian and 1 Pacific. Despite this, Africa and the Pacific were still represented by only 2 and 1 languages respectively, which is not desired but necessary due to the time frame of this paper, since finding fitting languages/language families would be too time consuming considering that the Niger-Congo family covers large parts of the African continent and the Austronesian and Trans-New Guinean families cover the Pacific region.

Table 6: Showing the geographical distribution of the five groups of languages containing less than 93 languages and the replacement languages.

Language Group	**South America**	**North America**	**Europe**	**Africa**	**Asia**	**Pacific**
<7	Ayoreo		Basque		Korean	
7-20	Epena	Karok	Scando-Romani			
21-44	Yaminahua	Tlingit	Estonian	Nama	White Hmong	
45-67	Wapishana Aché Imbabura Quechua	Nahuatl	Turkish			
7-20				Seychelles Creole	Thai	
Replacement languages	Catuquina	Inuktitut	Georgian		Ainu	Tok Pisin
		Cheyenne	Archi		Ket	
		Zinacantán Tzotzil			Japanese	
Total	7	5	6	2	6	1

For families represented by several languages, the choice of languages were spread over several branches, but also areally as much as possible (depending on available data), Table 7

and Figure 4. The Afro-Asiatic languages were represented by the Chadic, Semitic, Cushitic and Berber branches, both of the 2 Australian languages belong to the Pama-Nyungan branch due to lack of sources of the remaining branches, and similarly, the Austro-Asiatic languages are both Mon-Khmer, though belonging to different subgroups, Viet-Muong and Aslian. The Austronesian languages were divided into 9 Malayo-Polynesian languages (being by far the largest branch) and 3 Formosan languages in order to have more branches represented; Western Plains, East Formosan and Atayalic. The Indo-European languages were represented by the Slavic, Germanic, Celtic, Armenian and Iranian branches. The large Niger-Congo family was represented by 9 languages from the Volta-Congo subgroup, and more specifically 5 Narrow Bantu, 1 Kainji, 2 Kwa and 1 Defoid. In addition 2 languages belonged to the Mande subgroup and 1 to the Atlantic, Northern, Senegambian, in contrast to the 11 aforementioned languages which all being the Atlantic-Congo subgroup. The Nilo-Saharan languages were represented by Eastern Sudanic and Saharan branches, the Oto-Manguean by the Western and Eastern branches, the Sino-Tibetan language by 1 belonging to the Chinese branch and 4 to the Tibeto-Burman and the Trans-New Guinea "family" or group was only represented by the West, Timor-Alor-Pantar branch.

This yields 75 genetically and areally spread languages, representing 37 separate language families/isolates.

Table 7: The final sample of languages, including language families, the groups of languages containing less than 93 languages and replacement languages. Color coding relating to the positions of the sampled languages in Figure 4.

Language family	% of the world's languages	Sample of 75 languages (%)	Final sample (adjustment)	Languages in the sample
Afro-Asiatic	5.1	4 (3.8)	4	Hausa, Hebrew, Iraqw, Tarifiyt Berber
Australian	2.2	2 (1.7)	2	Gurindji, Warlpiri
Austro-Asiatic	2.5	2 (1.8)	2	Vietnamese, Cheq Wong
Austronesian	17.8	13 (13.4)	12 (-1)	Hawaiian, Tongan, Rotuman, Malagasy, Takia, Cebuano, South Efate, Tetum, Malay, Thao, Kavalan, Seediq
Indo-European	6.2	5 (4.6)	5	Czech, German, Breton, Armenian, Persian
Niger-Congo	21.9	16 (16.4)	12 (-4)	Swahili, Sesotho, Zulu, Nyanja, C'Lela, Akan, Ewe, Kinyarwanda, Yoruba, Bambara, Wolof, Mandinka
Nilo-Saharan	2.9	2 (2.2)	2	Ghulfan, Kanuri
Oto-Manguean	2.5	2 (2.5)	2	Otomi, Chatino
Sino-Tibetan	6.4	5 (4.8)	5	Mandarin, Manange, Tibetan, Mikir, Akha
Trans-New Guinea	6.9	5 (5.2)	1 (-4)	Fataluku
<7 languages/family, Isolates, Unclassified	2.2	3 (2.8)	3	Ayoreo, Basque, Korean
7-20 languages/family	3.7	3 (2.6)	3	Epena, Karok, Scando-Romani
21-44 languages/family	3.5	5 (4.9)	5	Yaminahua, Estonian, Tlingit, White Hmong, Nama
45-67 languages/family	6.5	5 (4.9)	5	Aché, Nahuatl, Wapishana, Imbabura Quechua, Turkish
68-92 languages/family	6.5	3 (2.7)	2 (-1)	Seychelles Creole, Thai
Replacement languages	-	-	10 (+10)	Catuquina, Ket, Japanese, Archi, Tok Pisin, Ainu, Inuktitut, Cheyenne, Zinacantán Tzotzil, Georgian
Total	100 %	75	75	

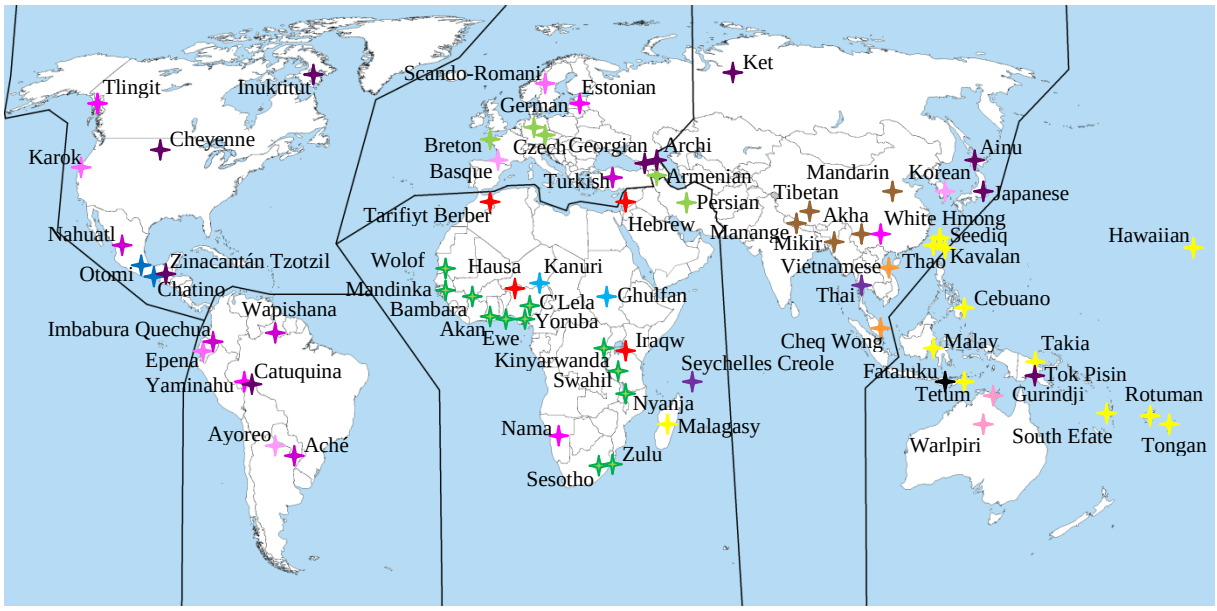

Figure 3: The positions of the sampled languages, color coded according to language family/group, see Table 7, and the six geographical areas used as guidelines in the areal distribution of the languages.

3.2 Concept Sampling

Since phonosemantics is a cross-linguistic phenomenon, this investigation is simplified by using universal concepts or concepts with universal tendencies as the object of study. The goal is that the chosen concepts are at least be present in the majority of the world's languages, as well as being binary pairs. There are many concepts showing close semantic relationships as well as being very fundamental whilst being more than simply a pair, such as colors and three-way deictic systems. These are however excluded in this study for methodological reasons.

In Goddard & Wierzbicka (2002) six oppositional pairs are presented, as well as three concepts which can easily be made into pairs; THIS, MUCH/MANY and HERE if the counterparts THAT, FEW and THERE are added. There were several concepts of the list of semantic primes which might have suited this investigation, though were ultimately excluded; VERY and MORE could both potentially be contrasted with LESS, however these concepts can be used as intensifiers for both poles of oppositional pairs e.g. less good can signify 'bad'. SOME, SOMEONE and SOMETHING/THING could be either contrasted with ALL or NOTHING, hence deciding on which of these words to use is rather difficult. TRUE ought to be connected to FALSE, though the sense of 'right' and 'wrong', might blur the meanings of these concepts. NOW can be contrasted with THEN (non-present), however since BEFORE and AFTER is already present, these were judged as the more suited time-related pair for this study. INSIDE and OUTSIDE are very clear though due to the often arbitrary use of

prepositions i.e. being on the train while actually sitting inside of it, and the related but different notions of something 'coming out of' or going 'into something' , these were excluded. SAME and OTHER could arguably be grouped, though OTHER does not have to signal the sense of 'different'. LIVE and DIE are clearly connected, though since they denote a process and a occurrence respectively, as well as the notion of 'being born' ought to be presented in most languages, these were excluded. ONE and TWO could possibly be considered contrasting something such as 'one' and 'more than one', though they are still far too gradient to be used. Lastly A LONG TIME and A SHORT TIME were excluded due to the former potentially being more widely used than the latter, since one often react to long durations, while short durations are overlooked. The remaining concepts were judged as having no suitable counterparts and were thus not included in this study.

If the same analysis is conducted on the 100-item Swadesh list (Swadesh 1971) we find six complete oppositional pairs; I-YOU (sg.), THIS-THAT, BIG-SMALL, WOMAN-MAN, WHITE-BLACK, HOT-COLD. Eight additional pairs can be created if FEW, SHORT, DAY, EMPTY, OLD, BAD and WET is added to MANY, LONG, NIGHT, FULL, NEW, GOOD, ROUND and DRY. Words with potential contrasts, though ultimately excluded include; WE could obviously be contrasted with YOU (pl.), though it is less clear than the singular counterparts since inclusive and exclusive WE are present in many of the world's languages. COME could be contrasted with GO, though GO has a very general sense in many languages, e.g. Tahitian *haere* 'go' and *haere mai* 'come'. SIT could perhaps be contrasted with STAND, though SIT could also be contrasted with some type of movement verb such as GO. GIVE could be contrasted with RECEIVE, GET but also with TAKE. LIVE, DIE, ONE, TWO, and ALL are explained above. Remaining words were excluded.

From the Swadesh subset studies confirmation for the full pair I and YOU (sg.) is found (Holman et al. 2008), as well as FULL, NEW and NIGHT, if EMPTY, OLD and DAY is added. In Pagel et al. (2013) THIS and THAT is confirmed, as well as MAN, MOTHER and BLACK if WOMAN, FATHER and WHITE is added. TO PULL is also featured as a very stable word which could be contrasted with 'to push' though the actual act of pulling and pushing differs, pulling is often accompanied by gripping and push simply by pushing, furthermore the notions might differ cross-linguistically, hence these were excluded.

The studies made by and featured in Willners (2001) only included English and Swedish antonyms (not present in the language sampling), despite this, they are used to reinforce the suitable concepts of investigation for this paper and as with the Semantic Primes and the Swadesh lists some words were excluded; BOTTOM-TOP and BEAUTIFUL-UGLY were excluded due to no occurrence in Junteson & Katz (1991). RIGHT, CORRECT-WRONG were excluded due to reasons explained above. DIRTY, SOILED-CLEAN were excluded due to potential cultural bias, being used in religious contexts. EASY-DIFFICULT, HARD could have a very relative sense e.g. the difference between hard to learn something and hard to carry something.

The final selection of suitable word pairs ended up as 28 pairs, Table 8; I-YOU, BIG-SMALL, GOOD-BAD, THIS-THAT and MUCH/MANY-FEW are represented by at least one of the contrasting concepts in the 100-item Swadesh list and in the Semantic Primes list. BEFORE-AFTER, ABOVE-BELOW, FAR-NEAR, WOMAN-MAN, WHITE-BLACK, HOT-COLD, HERE-THERE, LONG-SHORT, NIGHT-DAY, FULL-EMPTY, NEW-OLD, ROUND-FLAT and DRY-WET are represented by at least one of the contrasting concepts in the 100-item Swadesh list or in the Semantic Primes list. WIDE/BROAD-NARROW, THICK-THIN, SMOOTH-ROUGH, HEAVY-LIGHT, DARK-LIGHT, FAST-SLOW, HARD-SOFT, DEEP-SHALLOW, HIGH-LOW are added due to being concepts which are used to explain the world around us connected to the senses (usually and primarily sight) and are represented by at least one of the contrasting concepts in the studies presented by Willners (2001). Lastly the word pair MOTHER-FATHER is added due to probably existing in all the world' s languages and the fact that in most cases the linguistic forms of these concepts are rather similar independent of language, as well as MOTHER being present in Pagel et al. (2013). The most unmarked case/gender/number as possible is used which often results in nominative/absolutive, masculine, singular forms.

Table 8: Concepts investigated in this study and their occurrences in previous investigations presented in chapter 2.

Oppositional Pairs		(Goddard 2002)		(Swadesh 1955)		(Holman et al. 2008)		(Pagel et al. 2013)		Studies in (Willners 2001)		Sense concepts
I	YOU (sg.)	X	X	X	X	X	X					
BIG	SMALL	X	X	X	X					X	X	
GOOD	BAD	X	X	X						X	X	
THIS	THAT	X		X	X			X	X			
MUCH/ MANY	FEW	X		X								
BEFORE	AFTER	X	X									

ABOVE	BELOW	X	X									
FAR	NEAR	X	X							X	X	
WOMAN	MAN			X	X				X			
WHITE	BLACK			X	X				X	X	X	
HOT	COLD			X	X					X	X	
HERE	THERE	X										
LONG	SHORT			X						X	X	
NIGHT	DAY			X		X						
FULL	EMPTY			X		X				X	X	
NEW	OLD			X		X				X	X	
ROUND	FLAT			X								
DRY	WET			X						X	X	
WIDE/ BROAD	NARROW									X	X	X
THICK	THIN									X	X	X
SMOOTH	ROUGH									X	X	X
HEAVY	LIGHT (not HEAVY)									X	X	X
DARK	LIGHT (not DARK)									X	X	X
FAST	SLOW									X	X	X
HARD	SOFT									X	X	X
DEEP	SHALLOW									X	X	X
HIGH	LOW									X	X	X
MOTHER	FATHER							X				X

3.3 Sound Classification and Quantification

The linguistic forms were converted into IPA, Appendix A, but also simplified in some cases, similar to the ASJP transcription used by Brown et al. (2008) and Wichmann et al. (2010) (see section 2.3.4). Aspiration, various forms of co-articulations, tone and phoneme quantity were not taken into consideration (partly due to shortcomings in representation in orthographies). Ejectives were treated as voiceless plosives, implosives as voiced plosives, nasal vowels as oral vowels and in the same manner as consonant clusters were segmented, affricates and diphthongs were divided into their segments. Also the labio-velar approximant [w] is counted as a labial sound rather than velar.

The ASJP-transcription system of Brown et al. (2008) and Wichmann et al. (2010) could possibly be a better choice as a base for the sound classification and grouping for this thesis than the ones used due to the rigorous research behind it. However there are shortcoming in the ASJP system for the purpose of this study as well, e.g. stop and fricative varieties of any consonants, as well as rounded and unrounded variants of vowels were treated as the same

sound, which could hide interesting connections to certain concepts. Regardless of which of these systems is used, some aspects are lost.

Due to the mind not being as detailed as theoretical phonetics when it comes to the categorization of phonemes, combined with unreliability of orthographies and transcriptions, the phonemes found in the study were divided into larger groupings based on the findings of earlier investigations of the matter, including Ohala's frequency code, the sonority hierarchy, as well as overall occurrences of vowels, consonants, voiced and voiceless sounds.

3.3.1 Frequency

Voiceless and voiced consonants were divided into 4 groups; bilabial and labiodental (here: Labials); dental and alveolar (here: Alveolars); post-alveolar, retroflex, alveolo-palatal and palatal (here: Palatals); velar, uvular, pharyngeal, epiglottal and glottal (here: Velars). The cardinal vowels were grouped into 6 vowel groups; i-like, e-like, a-like, ə-like, o-like and u-like. These groups were then put on a scale of frequency based on the average frequency (Hz) of the second formant (F2) of the vowels, being the most varying formant (Lieberman & Blumstein 1988:171-184; Ladefoged 2005:40-48), and the average frequency of energy accumulation of the consonants (Ladefoged 2005:49-62), Table 9.

Table 9: Sound groups according to the F2 frequency of vowels and energy accumulation for consonants.

Voiceless				**Voiced**									
-	-	-	-	2500	2000+	1850-2300	1400-1800	1500-1750	1100-1450	1000-1500	500-1500	350-1050	200
Palatal	Alveo.	Velar	Labial	i-like	Palatal	e-like	Alveo.	a-like	ə-like	Velar	Labial	o-like	u-like
				i		y, e, ɛ, ø		æ, i, a, œ	ə, œ, ɐ, ʉ, ʌ			ɤ, ɑ, ɒ, ɯ, ɔ, o	u

3.3.2 Sonority

This grouping is made as an attempt to capture possible sound-meaning correlations based on sonority, based on Hogg & McCully (1987:32-33), e.g. texture-related concepts, Table 10. The vowel groups from Frequency were kept the same, since the manner of their articulation based on frequency were judged to be adequately fitting for this purpose as well. Rounded vowels were judged less sonorous than their unrounded counterparts due to them being more obstructed and [ə] was judged more sonorous than the e-group due to the tongue being in relaxed position. Consonants were divided into; Approximants, Trills/Tap/Flap (grouped into one due orthographies sometimes being unclear in disguising between [ɾ] and [r] etc.), Laterals, Nasals, Fricatives and Plosives. Fricatives and Plosives were not divided into voiced

and voiceless since the difference in sonority is rather small, hence it seems more beneficial to group these together for the purpose of this paper.

Table 10: Sound groups according to sonority, ranging from the most sonorous (a-like) to the least sonorous (Plosive).

a-like	ə-like	e-like	o-like	i-like	u-like	Approximant	Trill/Tap/Flap	Lateral	Nasal	Fricative	Plosive

more sonorous ⟷ less sonorous

3.3.3 Combination

The same grouping of sounds as according to Frequency was repeated with the difference of including one more dimension, the aspects of continuousness and abruptness, i.e. the sonority hierarchy. This is done by dividing the sounds into four groups, vowels, again i-like, e-like, a-like, ə-like, o-like and u-like (V); semi-vowels, approximants, nasals and laterals (A); fricatives, dorsal trills and voiceless sonorants (F); plosive, apical trills, taps and flap (P), meaning that the vowels are kept the same, while the voiceless consonants have two types of groups (F and P) for each place of articulation and the voiced consonants have three (F, P and A). The trills were treated this way due to the tendency of [r] being reduced to [ɾ] in many languages and contexts, as well as [ʁ] being realized as [ʁ] or [χ] or similar sounds, Table 11.

This grouping might capture e.g. phonesthemes, which are built up by consonants clusters, sometimes possibly cross-linguistically, by combining the place of articulation with the manner of articulation.

Table 11: The 26 sound groups of the Combination-grouping, combining the Frequency-grouping with aspects of the Sonority-grouping.

	Voiceless				**Voiced**									
	Palatal	Alveo.	Velar	Labial	i-like	Palatal	e-like	Alveo.	a-like	ə-like	Velar	Labial	o-like	u-like
V	-	-	-	-	9	-	13	-	17	18	-	-	25	26
A	-	-	-	-	-	10	-	14	-	-	19	22	-	-
F	1	3	5	7	-	11	-	15	-	-	20	23	-	-
P	2	4	6	8	-	12	-	16	-	-	21	24	-	-

3.3.4 General

In the General grouping the sounds were dividing into very general groups; total voiceless and voiced sounds, as well as into total consonants and vowels, in order to see potential differences on a very general phonetic level, Table 12.

Table 12: The large sound groups including very general phonetic traits such as voicing and types of speech sounds.

Voicing		Type of Speech sound	
Voiceless Sounds	Voiced Sounds	Consonants	Vowels

3.4 Data Sources

The language data was collected from various reliable sources including The Intercontinental Dictionary Series, World Loanword Database, Free Personal Pronoun System database, Swadesh lists found through the Rosetta Project, dictionaries (including online[2]), as well as first hand sources of scholars and native speakers, see Language Sources under References. In cases when more than one word is given per concept, the first form will be used since they are probably most commonly used, provided that the author does not have prior knowledge of the language in question. In many cases all linguistic forms for all concepts were not found, however in 38 of the 56 concepts over 70 linguistic forms were found, and the gaps of the remaining 18 concepts were quite spread out among the languages.

3.5 Analyses

All occurrences for each of the sound groups in each grouping and each concept were examined through two analyses.

3.5.1 Related Phoneme Distribution

The four sound groupings explained above, *Frequency*, *Sonority*, *Combination* and *General*, as well as a fifth grouping, *All*, containing all sound groups of the four described groupings, were analyzed by means of cluster analyses creating biplots for each of the five groupings using the statistical computing and graphics software R (see Johnson 2008). For *All*, the vowels were only counted once since as opposed to the consonants, the vowel groups were identical in *Frequency*, *Sonority* and *Combination*, otherwise all other sound groups were included as they were. The cluster analyses creating biplots were conducted by measuring how similar the phoneme-distribution was among the investigated concepts (based on the calculated percentages of each occurrence of each sound group for each concept, e.g. the number of nasals in BIG or the number of Voiced Alveolars in BELOW). All phonemic parameters were then projected onto the two-dimensional biplots, indicating relative closeness in sound composition between all concepts. Concepts that are located closely together indicate similar phoneme-distributions, while occurrences far away from each other indicate the usage of different phonemes for the concepts in question. This shows whether semantically related concepts also correlate sounds-wise.

[2] It has to be mentioned that using online dictionaries always carry with it a small risk of unreliable data, though since the concepts used in this thesis are very fundamental in nature, the words extracted ought to be reliable.

By using the different sound groupings, various phonetic traits are captured allowing many different kinds of potential sound-meaning associations to be visible, i.e. different clustering of concepts depending on which of sound groupings that are investigated.

3.5.2 Deviation from Average Phoneme Distribution

Lastly the phoneme distributions' deviation from the normal was calculated (over- and underrepresentation) using the average distributions found in Brown et al. (2008) and Wichmann et al. (2010), converted from their so called ASJP-transcription system to the system used in this thesis.

Two thresholds of deviation was used, first 50 % both over- and underrepresentation, and secondly 100 % overrepresentation, while the underrepresentation was 75 %, since 100 % underrepresentation would only apply to phonemes with no occurrence at all.

These values were then used as guidelines for which phonemes represented each concept, as well as which phonemes were representing the concept, but also lacking to a higher degree, which is another important factor for the makeup of the concepts.

4. Results

4.1 Significant Concepts

Concepts with diverging phoneme distributions causing them to occur outside the 0.05/-0.05-radius from the center of the biplots were judged to be significant[3], this radius is also the boundary of the non-significant concepts found in Wichmann et al. (2010), presented below. The distribution of the concepts in the biplots shows that within the 0.05/-0.05 boundary rather apparent group of various concepts are found, constituting concepts with normal phoneme distribution, while remaining concepts form outlying clusters in all directions. The locations of the concepts within the plots can then be correlated with the deviation from average phoneme distribution for each sound group (presented in section 4.5 below), which indicates which over- and underrepresentations of sound groups the different concepts have in common.

The concepts outside of the 0.05-radius ought to be different from the average phoneme distribution enough to consider them interesting and significant for analyses, Table 13. 15 of

[3] Note that *significant* does not in this case mean *statistically signnificant* in terms of p-values, but *significantly distinct phoneme distributions* to be judged as non-random.

the 56 featured concepts did not occur outside of the 0.05-radius in any of the biplots; GOOD, BAD, MUCH/MANY, BEFORE, AFTER, BELOW, FAR, NIGHT, NEW, DRY, THIN, HEAVY, FAST, HIGH, LOW. Of the remaining concepts 13 occurred in all of the sound groupings; MOTHER, I, FATHER, WIDE/BROAD, ABOVE, MAN, THIS, NARROW, SMALL, ROUGH, SHORT, BLACK, FLAT. Four of the concepts occurred outside of the radius in four of the groupings; SHALLOW, DARK, WHITE, YOU. 12 in three of the groupings; WOMAN, THERE, LONG, DEEP, SOFT, ROUND, HARD, NEAR, HOT, WET, FEW, DAY. Six in two of the groupings; THAT, SLOW, LIGHT (not HEAVY), COLD, FULL, EMPTY. Seven in one grouping; SMOOTH, LIGHT (not DARK), OLD, HERE, BELOW, THICK, BIG. LIGHT (not DARK) and OLD only passed the 0.05-raidus in the All-plot, possibly showing stronger crystallization of the results when including more sound groups into one analysis.

Table 13: The concepts occurring outside of the 0.05-radius in the five different sound groupings.

	Frequency	Sonority	Combination	General	All	
MOTHER , I, FATHER , WIDE/BROAD, ABOVE, MAN, THIS, NARROW, SMALL, ROUGH, SHORT, BLACK, FLAT	X	X	X	X	X	5
SHALLOW, DARK	X	X	X		X	4
WHITE	X	X	X	X		4
YOU		X	X	X	X	4
WOMAN		X		X	X	3
THERE		X	X		X	3
LONG, DEEP	X				X	3
SOFT, ROUND	X	X			X	3
HARD, NEAR		X		X	X	3
HOT, WET, FEW, DAY	X			X	X	3
THAT				X	X	2
SLOW		X		X		2
LIGHT (not HEAVY)	X	X				2
COLD, FULL, EMPTY				X	X	2
SMOOTH, LIGHT (not DARK), OLD					X	1
HERE, BELOW				X		1
THICK, BIG		X				1

4.2 Significant Clusters

Even more interesting are the clusters forming outside of the 0.05-radius in each of the different biplots, and especially those which contain concepts with similar semantic content. If the significant concepts are classified and summarized according to semantic domain, independently of how they are clustered, 10 domains and one lone concept were found; Small,

Intense Vision-Touch (which all are vision-touch-related), Large, Organic, Horizontal-Vertical Distance, Deictic, Containment, Gender, Parent and Diurnal, as well as OLD.

All domains were represented in all plots except for Large and Organic which were not significant in the Combination-plot; Organic was not significant in the General-plot; Containment was lacking in the Frequency-, Sonority- and Combination-plots; Diurnal in the Sonority- and Combination-plots and lastly OLD only present in the All-plot. However, even when these domains were not significant, the overall distribution of the domains in the plots show similar relative patterns also within the 0.05-radius.

Generally these domains correspond to how the concepts are grouped in the biplots with only a few exceptions, Table 14. In the Frequency-plot LIGHT (not HEAVY) is not grouped with the rest of the Intense Vision-Touch concepts. In the Sonority-plot BIG is not found among the Large concepts, SLOW is not found among the Organic concepts, and WHITE and LIGHT (not HEAVY) are not found among the Intense Vision-Touch concepts. And lastly SLOW is not found among the Organic concepts in the General-plot. THERE could be classed as both Deictic and Horizontal-Vertical Distance, though it only occurs together with Horizontal-Vertical Distance concepts. Furthermore, the actual semantic clusters showed very similar relative positions to each other regardless of type of sound grouping.

Table 14: Significant concepts divided according to semantic domain.

Domain	Frequency	Sonority	Combination	General	All
Small	SHORT, SMALL, NARROW, FEW	SHORT, SMALL, NARROW, NEAR	SHORT, SMALL, NARROW	SHORT, SMALL, NARROW, FEW, NEAR	SHORT, SMALL, NARROW, FEW, NEAR
Intense Vision-Touch	DARK, BLACK, WHITE, HOT, WET, LIGHT (not HEAVY), ROUGH	DARK, BLACK, WHITE, LIGHT (not HEAVY), ROUGH, HARD	DARK, BLACK, WHITE, ROUGH	BLACK, HOT, WET, COLD, WHITE, ROUGH, HARD	DARK, BLACK, HOT, COLD, WET , ROUGH, HARD
Large	DEEP, LONG	BIG		LONG, DEEP	DEEP, LONG
Organic	ROUND	SOFT, THICK, ROUND, SLOW			SOFT, ROUND, SMOOTH, SLOW
Horizontal-Vertical Distance	FLAT, SHALLOW, WIDE/BROAD, ABOVE	FLAT, SHALLOW, WIDE/BROAD, ABOVE	FLAT, SHALLOW, WIDE/BROAD, ABOVE	ABOVE, FLAT, WIDE/BROAD, BELOW	SHALLOW, WIDE/BROAD, FLAT, ABOVE
Deictic	I, THIS	I, THIS, YOU, THERE	I, THIS, YOU, THERE	I, THIS, YOU, THAT, HERE	I, THIS, YOU, THAT, THERE
Containment				EMPTY, FULL	FULL, EMPTY
Gender	MAN	WOMAN, MAN	MAN	WOMAN, MAN	WOMAN, MAN
Parent	MOTHER, FATHER	MOTHER, FATHER	MOTHER, FATHER	MOTHER, FATHER	MOTHER, FATHER
Diurnal	DAY			DAY	DAY, LIGHT (not DARK)
OLD					OLD

4.3 The Biplots

Below the five biplots content and makeup is presented for the five types of sound groupings; Frequency, Sonority, Combination, General and All, explained in section 3.3 and 3.5. Clusters of semantically related concepts outside of the 0.05-radius were highlighted by colors. If a semantic domain were found to have at least one member outside of the 0.05-radius, the related concepts within the 0.05-radius were also highlighted for reference and interesting connections. These significant-by-association concepts were marked with an asterisk when described in text.

The Frequency-plot shows, Figure 5, that the Parent concept, MOTHER and FATHER, are both located far away from the center, each in its own direction. Closer to the center three major clusters are found. The first containing the deictic concepts THIS, THAT*, HERE*, YOU*; large-category concepts DEEP, LONG, FAR*, BIG*; organic tactile ROUND, HEAVY*, SOFT*, THICK*, SMOOTH* and lastly DAY. The second cluster contain horizontal and vertical distance concepts WIDE/BROAD, SHALLOW, FLAT, ABOVE, THERE*, HIGH* as well as I, LIGHT (not HEAVY) and MAN. Finally the last is a very complex cluster consisting of six major semantic categories; small-category concept SHORT, SMALL, NARROW, FEW; light/color concepts DARK, BLACK, WHITE; temperature concepts HOT, COLD*; moisture concepts WET, DRY*; inorganic tactile concepts ROUGH, HARD*.

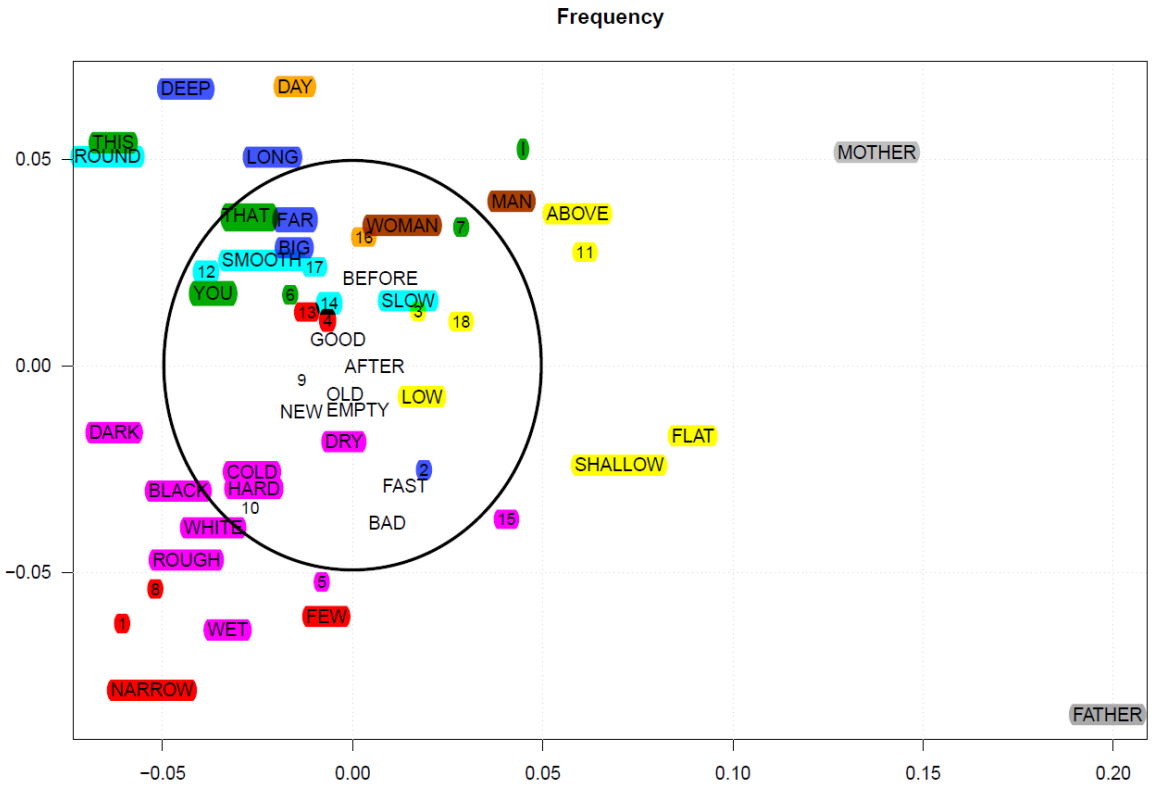

Figure 4: Biplot according to Frequency. Some numbers adjusted very slightly in order to be readable. Concepts near the origin have average phoneme distribution, indicated by the black circle. Concepts outside the red circle deviate from the average distribution. Clusters of deviating semantically related concepts are indicated by the blue ellipses. Number legend: 1: SMALL, 2: MUCH/MANY, 3: BELOW, 4: NEAR, 5: HOT, 6: HERE, 7: THERE, 8: SHORT, 9: NIGHT, 10: FULL, 11: WIDE/BROAD, 12: THICK, 13: THIN, 14: HEAVY, 15: LIGHT (not HEAVY), 16: LIGHT (not DARK),17: SOFT, 18: HIGH. Colors legend: Small - red, Intense Vision-Touch - magenta, Large - dark blue, Organic - light blue, Horizontal-Vertical Distance - yellow, Deictic - dark green, Containment - light green, Gender - brown, Parent - grey, Diurnal - orange, OLD - light purple.

In the Sonority-plot, Figure 6, MOTHER and FATHER are once again found far away from the center, though here this applies for I as well. A horizontal and vertical distance concept cluster is found above the center, consisting of FLAT, SHALLOW, WIDE/BROAD, THERE, ABOVE, as well as LIGHT (not HEAVY). Close by SLOW is found, as well as the gender concepts WOMAN, MAN and further towards the bottom the deictic concepts YOU, THIS, THAT*. The bottom right section is very complex but can be divided into two clusters. The first one consisting of organic tactile concepts SOFT, THICK, ROUND, FULL*, HEAVY*, SMOOTH*; the large-category concepts LONG*, FAR*, DEEP*. In the second small-category concepts NEAR, SMALL, SHORT, NARROW are found; inorganic tactile concepts HARD, ROUGH; as well as BIG and WHITE. Light/color related concept DARK, BLACK, NIGHT* as well as WET* and COLD* are founded shared between the two previously mentioned clusters.

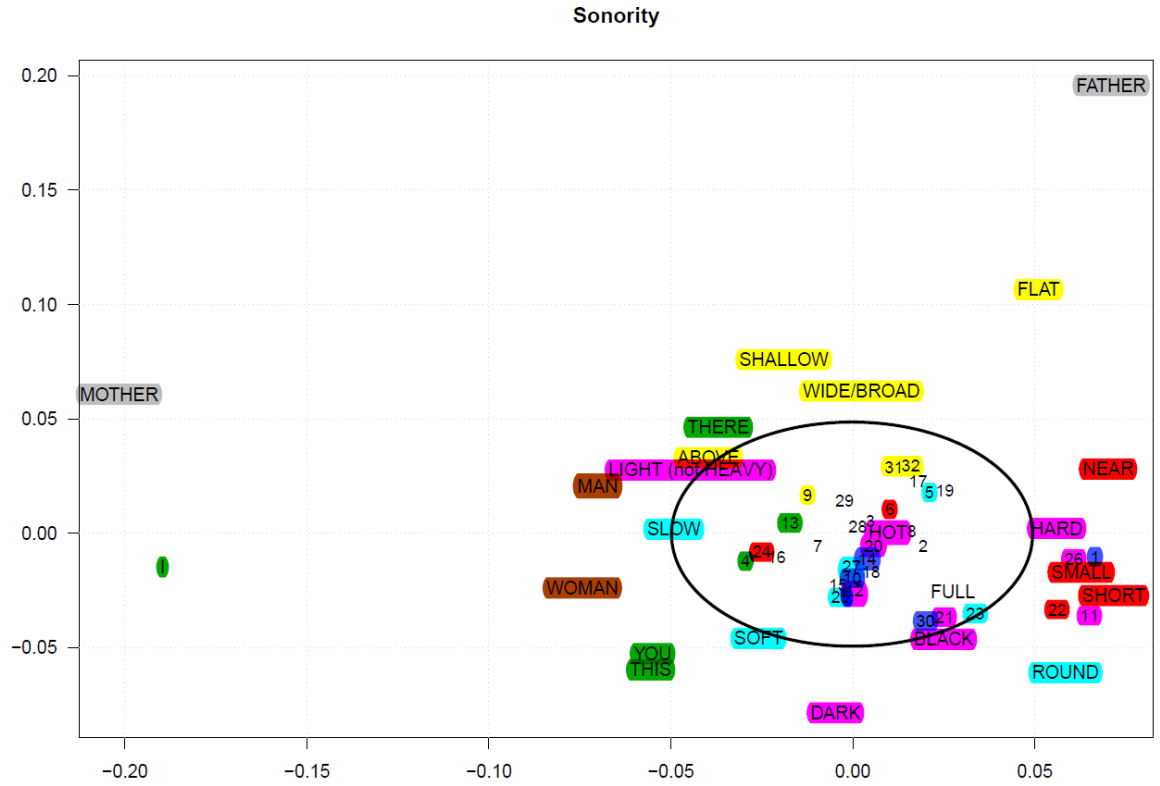

Figure 5: Biplot according to Sonority. Some numbers adjusted very slightly in order to be readable. Concepts near the origin have average phoneme distribution, indicated by the black circle. Concepts outside the red circle deviate from the average distribution. Clusters of deviating semantically related concepts are indicated by the blue ellipses. Number legend: 1: BIG, 2: GOOD, 3: BAD, 4: THAT, 5: MUCH/MANY, 6: FEW, 7: BEFORE, 8: AFTER, 9: BELOW, 10: FAR, 11: WHITE, 12: COLD, 13: HERE, 14: LONG, 15: NIGHT, 16: DAY, 17: EMPTY, 18: NEW, 19: OLD, 20: DRY, 21: WET, 22: NARROW, 23: THICK, 24: THIN, 25: SMOOTH, 26: ROUGH, 27: HEAVY, 28: LIGHT (not DARK), 29: FAST, 30: DEEP, 31: HIGH, 32: LOW. Colors legend: Small - red, Intense Vision-Touch - magenta, Large - dark blue, Organic - light blue, Horizontal-Vertical Distance - yellow, Deictic - dark green, Containment - light green, Gender - brown, Parent - grey, Diurnal - orange, OLD - light purple.

Again MOTHER, FATHER and I are found far away from the center in the Combination-plot, Figure 7. And again a Horizontal-Vertical Distance cluster consisting of FLAT, SHALLOW, WIDE/BROAD, THERE, ABOVE as well as LIGHT (not HEAVY)* and MAN is found. Several clusterings are located on the right side, quite clearly divided. From the top; small-category concepts NARROW, SHORT, SMALL, NEAR, FEW; inorganic tactile concepts ROUGH, HARD*; temperature concepts HOT*, COLD*; light/color concepts BLACK, DARK, WHITE; moisture concepts WET, DRY. Further down organic tactile concepts ROUND, HEAVY*, THICK*, SOFT*, SMOOTH* are located. At the very bottom the deictic concepts YOU, THIS, THAT* are found.

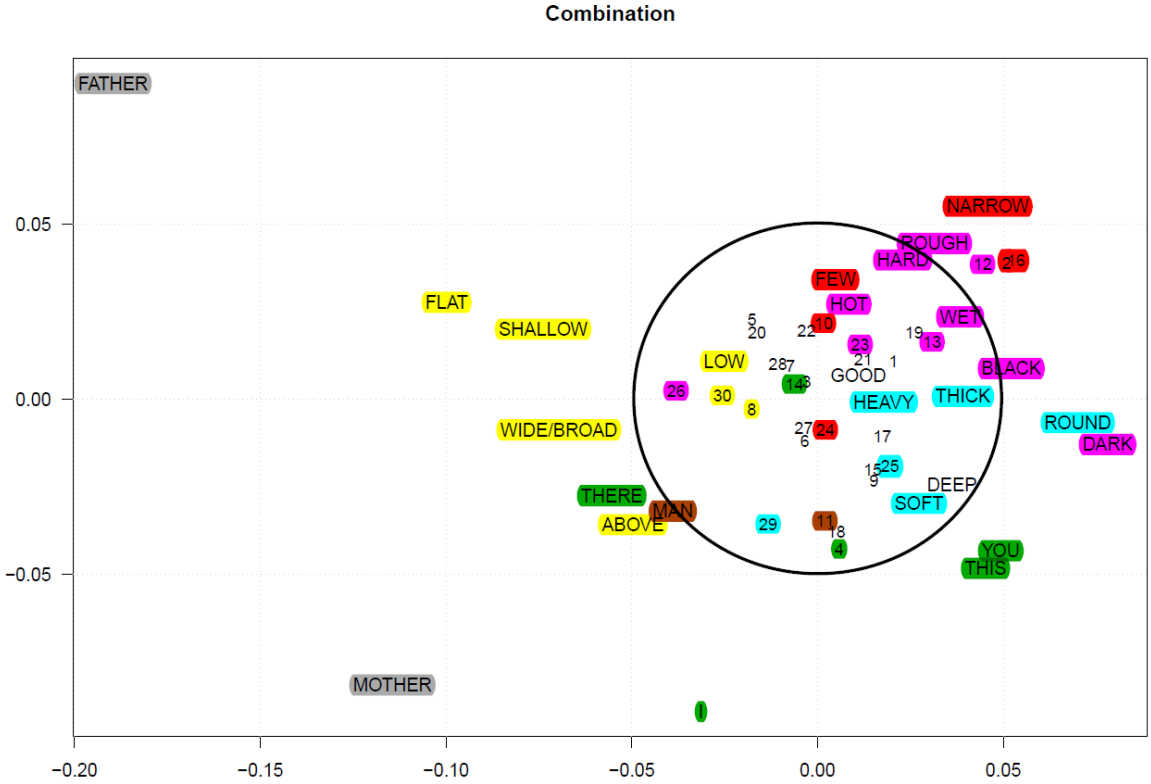

Figure 6: Biplot according to Combination. Some numbers adjusted very slightly in order to be readable. Concepts near the origin have average phoneme distribution, indicated by the black circle. Concepts outside the red circle deviate from the average distribution. Clusters of deviating semantically related concepts are indicated by the blue ellipses. Number legend: 1: BIG, 2: SMALL, 3: BAD, 4: THAT, 5: MUCH/MANY, 6: BEFORE, 7: AFTER, 8: BELOW, 9: FAR, 10: NEAR, 11: WOMAN, 12: WHITE, 13: COLD, 14: HERE, 15: LONG, 16: SHORT, 17: NIGHT, 18: DAY, 19: FULL, 20: EMPTY, 21: NEW, 22: OLD, 23: DRY, 24: THIN, 25: SMOOTH, 26: LIGHT (not HEAVY), 27: LIGHT (not DARK),28: FAST, 29: SLOW, 30: HIGH. Colors legend: Small - red, Intense Vision-Touch - magenta, Large - dark blue, Organic - light blue, Horizontal-Vertical Distance - yellow, Deictic - dark green, Containment - light green, Gender - brown, Parent - grey, Diurnal - orange, OLD - light purple.

In the General-plot, Figure 8, MOTHER and I are found far from the center, though FATHER is here located just above the center. Rather close to the location of I, a distinct Deictic concept cluster is found, consisting of YOU, THIS, THAT, HERE, THERE*. Below a vertical concept cluster consisting of ABOVE, BELOW, HIGH* is located, and further down a gender concept cluster consisting of WOMAN, MAN. Towards the bottom a cluster of large-category concepts LONG, DEEP, FAR*, BIG*; the horizontal concept WIDE/BROAD and the Diurnal concepts DAY and LIGHT (not DARK)* is found. To the right a complex cluster is found again mostly consisting of small category concepts SMALL, NEAR, SHORT, NARROW, FEW, but also of horizontal concepts FLAT, SHALLOW*, moisture concept WET, DRY*; temperature concepts HOT, COLD; light/color concepts BLACK, WHITE, DARK*; inorganic tactile concepts ROUGH, HARD; and containment concepts EMPTY, FULL.

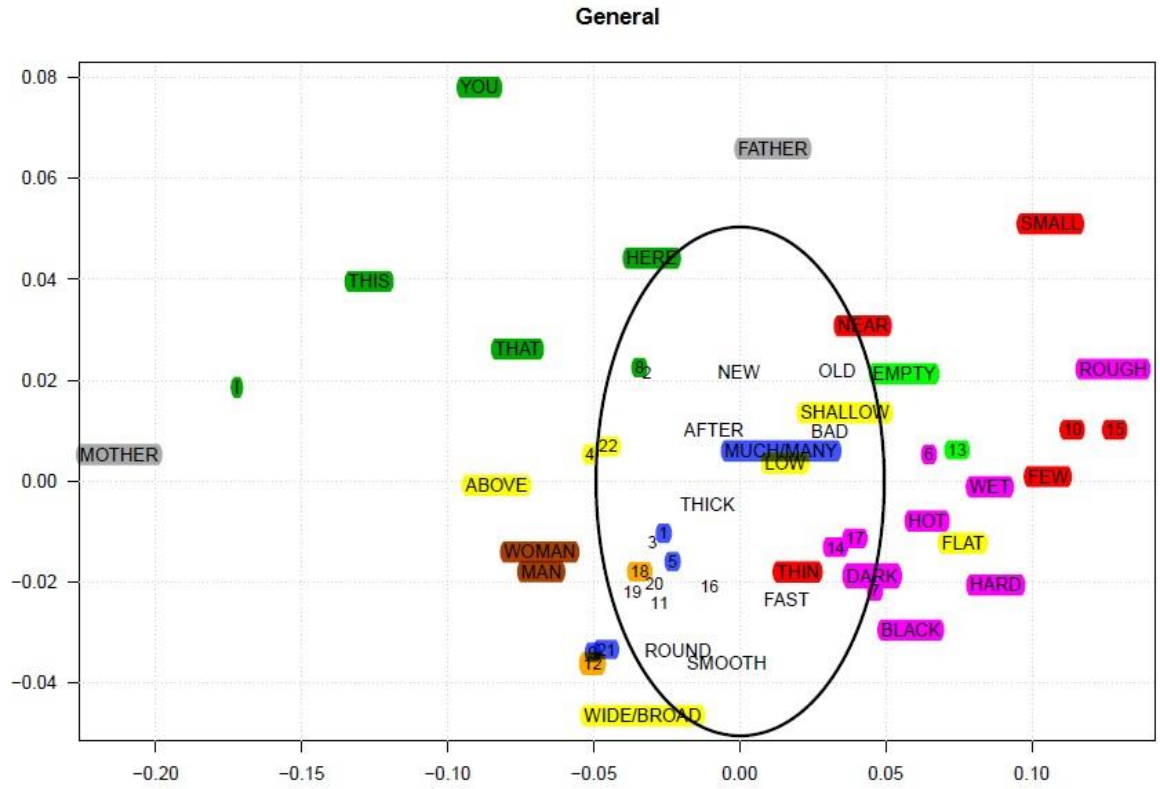

Figure 7: Biplot according to General. Some numbers adjusted very slightly in order to be readable. Concepts near the origin have average phoneme distribution, indicated by the black circle. Concepts outside the red circle deviate from the average distribution. Clusters of deviating semantically related concepts are indicated by the blue ellipses. Number legend: 1: BIG, 2: GOOD, 3: BEFORE, 4: BELOW, 5: FAR, 6: WHITE, 7: COLD, 8: THERE, 9: LONG, 10: SHORT, 11: NIGHT, 12: DAY, 13: FULL, 14: DRY, 15: NARROW, 16: HEAVY, 17: LIGHT (not HEAVY), 18: LIGHT (not DARK),19: SLOW, 20: SOFT, 21: DEEP, 22: HIGH. Colors legend: Small - red, Intense Vision-Touch - magenta, Large - dark blue, Organic - light blue, Horizontal-Vertical Distance - yellow, Deictic - dark green, Containment - light green, Gender - brown, Parent - grey, Diurnal - orange, OLD - light purple.

When all sound groupings are put together in the All-plot, Figure 9, MOTHER, FATHER and I are located very far away from the center. Above the center horizontal and vertical concept FLAT, SHALLOW, WIDE/BROAD, ABOVE, BELOW and the deictic concept THERE are found. Towards the left the gender concept WOMAN, MAN and the deictic concepts YOU, THIS, THAT, THERE, HERE* are located. At the bottom large-category concepts LONG, DEEP, FAR*, BIG*; organic tactile concepts SOFT, ROUND, SMOOTH, SLOW, HEAVY*, THICK* are located, as well as the Diurnal concepts DAY and LIGHT (not DARK)*. Towards the right small-category concepts NEAR, FEW, NARROW, SMALL, SHORT; inorganic tactile concepts HARD, ROUGH; light/color concepts DARK, BLACK; temperature concepts HOT, COLD; moisture concepts WET, DRY*; containment concepts FULL, EMPTY, as well as OLD is found.

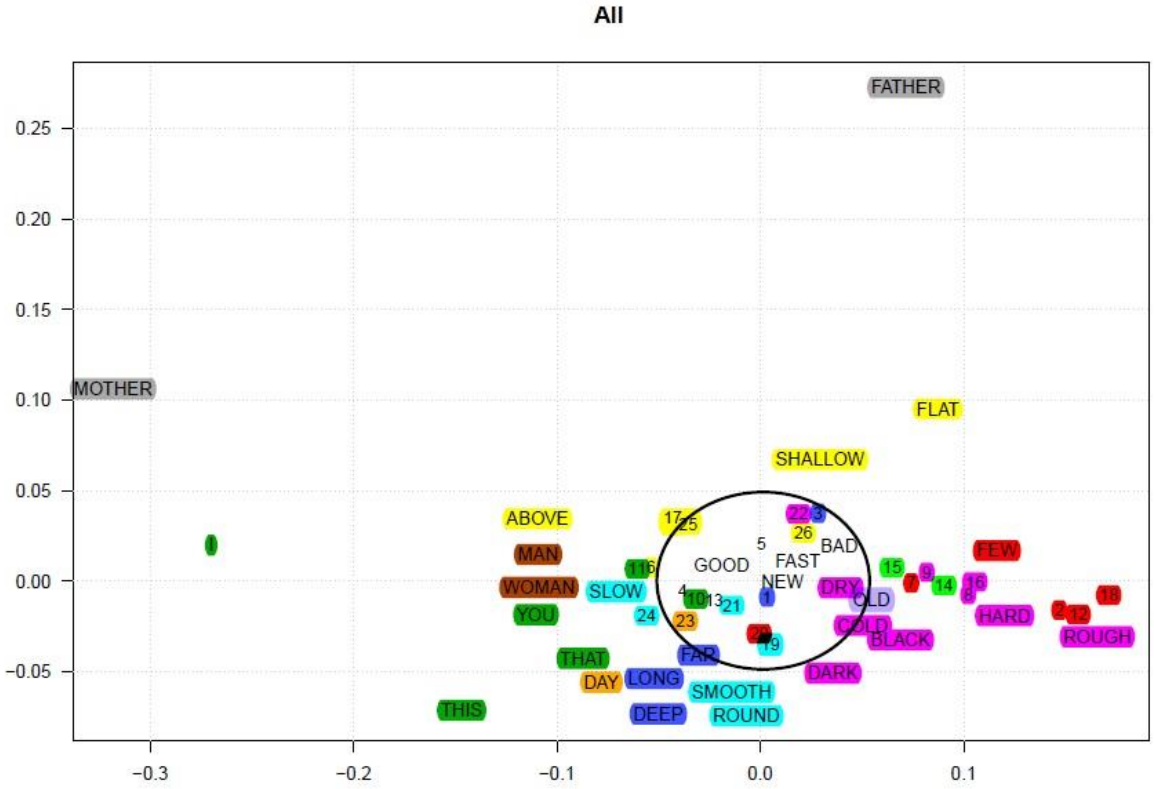

Figure 8: Biplot according to All. Some numbers adjusted very slightly in order to be readable. Concepts near the origin have average phoneme distribution, indicated by the black circle. Concepts outside the red circle deviate from the average distribution. Clusters of deviating semantically related concepts are indicated by the blue ellipses. Number legend: 1: BIG, 2: SMALL, 3: MUCH/MANY, 4: BEFORE, 5: AFTER, 6: BELOW, 7: NEAR, 8: WHITE, 9: HOT, 10: HERE, 11: THERE, 12: SHORT, 13: NIGHT, 14: FULL, 15: EMPTY, 16: WET, 17: WIDE/BROAD, 18: NARROW, 19: THICK, 20: THIN, 21: HEAVY, 22: LIGHT (not HEAVY), 23: LIGHT (not DARK), 24: SOFT, 25: HIGH, 26: LOW. Colors legend: Small - red, Intense Vision-Touch - magenta, Large - dark blue, Organic - light blue, Horizontal-Vertical Distance - yellow, Deictic - dark green, Containment - light green, Gender - brown, Parent - grey, Diurnal - orange, OLD - light purple.

4.4 Oppositional Relations

The relative positions of the concepts clusters in the biplots were found to be more or less the same, hence Figure 10 can be used as a summary of the Releated Phoneme Distribution results. The Frequency-, Sonority- and Combination-groupings were the most similar, basically only differing in orientation around the 0.0-0.0 point. The General-grouping gave the least defined results, obviously since the sound groups featured were very broad and the All- grouping gave the clearest results on account of having the most data.

Out of the 41 significant concepts only five significant concepts did not have a significant antipodes, OLD, DAY, WET, LIGHT (not HEAVY), SLOW, while the remaining 37 concepts all were included in some kind of semantic oppositional relationships. 10 pairs had their oppositional pairs located in separate clusters, MOTHER-FATHER, SOFT-HARD, SMOOTH-ROUGH, BIG-SMALL, LONG-SHORT, FLAT-ROUND, DEEP-SHALLOW,

DARK-LIGHT (not DARK), I-YOU, HERE-THERE, while five pairs within the same cluster, WOMAN-MAN, FULL-EMPTY, THIS-THAT, HOT-COLD, BLACK-WHITE. I is contrasted with YOU, though YOU is clustered together with THIS, THAT and HERE, which is in turn contrasted with THERE in yet another cluster, creating a three degree distinction in person/distance. And lastly, there is another three-way distinction between the Large (possibly together with Organic), Small (possibly together with Intense Vision-Touch) and Horizontal-Vertical Distance, with the difference of them not acting as grades on a scale, but all are contrasted with each other.

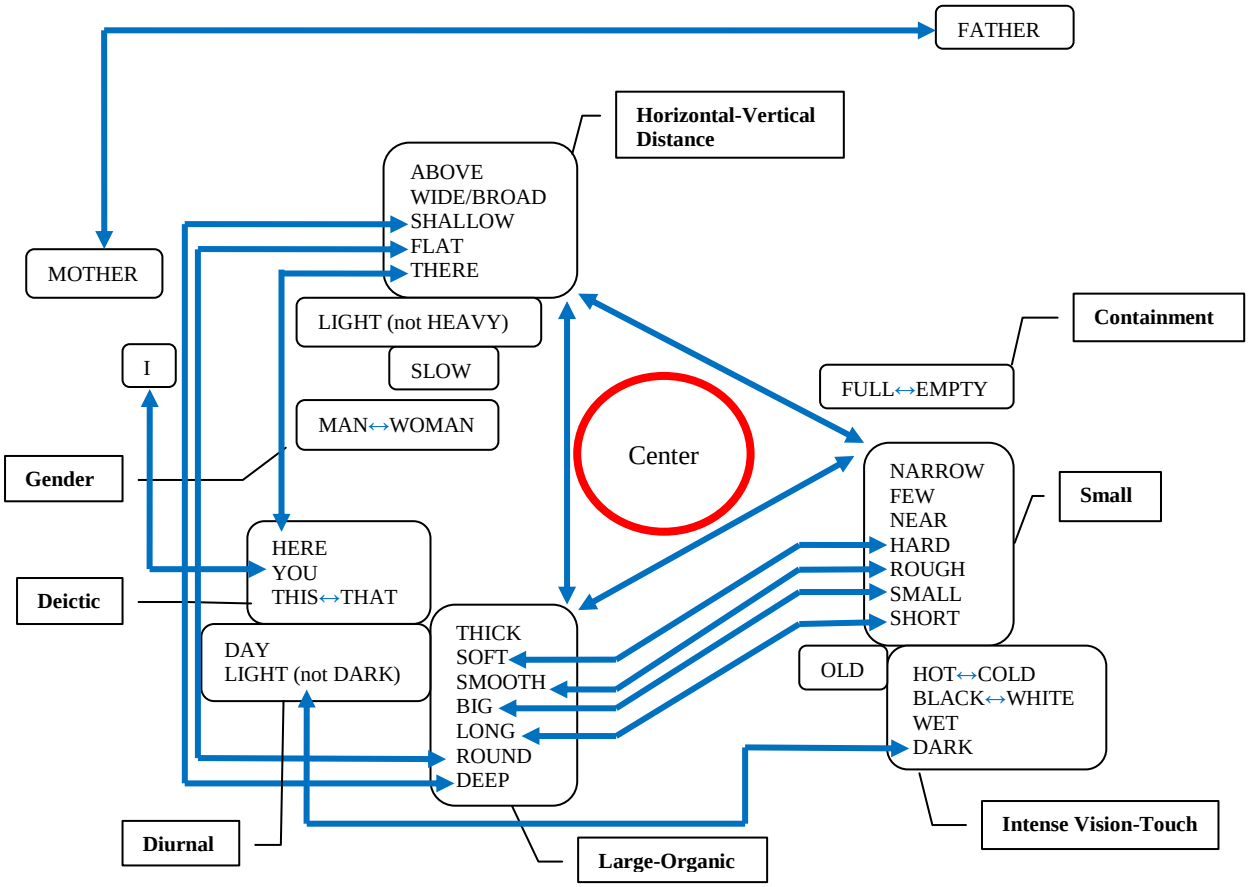

Figure 9: Simplified, though still accurate positions of the concepts judged significant of the five biplots, as well as relations between them. Rounded text boxes contains concepts of each semantic domain cluster and square text boxes shows the semantic domain name. Blue arrows shows oppositional relations between clusters, small blue arrows shows oppositional relations within clusters and the red circle indicates the center of the biplots, i.e. average phoneme distribution.

4.5 Deviation from Average

Analyzing the phoneme distributions found in the investigated concepts and comparing them with average phoneme distributions (by 50 % and 100/75 % over- and underrepresentation)

allows for interpretation of which sound groups are responsible for the results in the biplots, i.e. giving a defining sound makeup for each investigated concept[4].

Several sounds groups were found to be very unreliable including types of sounds which have generally low average occurrences. For the biplots this does not pose a large problem due to the very low occurrences, though when measuring in percentages, a slight error can seem like a large deviation, hence reliable sound groups were separated from unreliable sound groups. Reliable sound groups include all sound groups of the Sonority-, General- and Frequency-groupings, except for Voiceless Palatal and ə-like, Table 15. In addition to the two aforementioned, all sound groups of Combination were judged unreliable, though their results are presented in Appendix B. Voiceless Alveolar (judged as reliable) could also be judged as unreliable as it is never being underrepresented, at least not with a minimum factor of 50 %, hence it should be viewed with caution.

Table 15: Deviation of phoneme distribution compared to average of reliable sound groups, including all sounds groups of the Sonority-, General- and Frequency-groupings, except for Voiceless Palatal and ə-like. 50 % overrepresentation (+ in light green), 100 % overrepresentation (++ in green), 50 % underrepresentation (- in light red), 75 % underrepresentation (-- in red).

	i-like	e-like	a-like	o-like	u-like	Voiceless Alveolar	Voiceless Velar	Voiceless Labial	Voiced Palatal	Voiced Alveolar	Voiced Velar	Voiced Labial	Nasal	Plosive	Fricative	Approximant	Trill/Tap/Flap	Lateral	Voiceless Sounds	Voiced Sounds	Consonants	Vowels
I					+			--	+		+		++				-	--				
YOU					+			-	+		--						-	--				
BIG																	+	+				
SMALL						++						-										
GOOD																						
BAD						+									+							
THIS		+						-	++			-					-					
THAT						+		--	+			-						+				
MUCH/MANY						+																
FEW						++												-				
BEFORE																	+					
AFTER											+						+					
ABOVE	-																	+				
BELOW						+											+					
FAR															+			+				
NEAR																-		+				
WOMAN													+		+			-				
MAN																						
WHITE						+		+					-		+		++					
BLACK						+										-		+				
HOT						+										-		+				
COLD						++		-									+					
HERE					-				+			-			+			-				
THERE												-						+				

[4] Worth noting is that BIG does not have over representation of a-like or u-like phonemes and SMALL does not have over representation of i-like phonemes despite these sound-meaning corraltions are frequently found in experimental studies.

	i-like	e-like	a-like	o-like	u-like	Voiceless Alveolar	Voiceless Velar	Voiceless Labial	Voiced Palatal	Voiced Alveolar	Voiced Velar	Voiced Labial	Nasal	Plosive	Fricative	Approximant	Trill/Tap/Flap	Lateral	Voiceless Sounds	Voiced Sounds	Consonants	Vowels
LONG										+	++						+	++				
SHORT					+	++																
NIGHT					+	+												+				
DAY										+							+	++				
FULL						++		++							+	-						
EMPTY											+				+							
NEW						+									+		+					
OLD						+					-				+	-	++					
ROUND				+	+						+						++	+				
FLAT			+	-		+		++								-		++				
DRY						+					+				+		+					
WET						++									+	-						
WIDE/BROAD											+						+	++				
NARROW						++		+	-						+	-						
THICK						+										-	+	+				
THIN						+											+	+				
SMOOTH						+				+					+	-	++	++				
ROUGH						++						-	-		++	-	++					
HEAVY																-	+					
LIGHT (not HEAVY)				-		+									+							
DARK					+	+										-		+				
LIGHT (not DARK)											+						++	+				
FAST				-		+									+							
SLOW																		+				
HARD						++					+						++					
SOFT				+				+									+	+				
DEEP					+					+	+					-	++	++				
SHALLOW					-	+									+	-						
HIGH																						
LOW				-		+		+			-					-	-	+				
MOTHER				-			-	-				+	++					-	-			
FATHER			++		-	+	-	++		-	-		-	+		-		--				

5. Discussion

5.1 Methodological Issues

Errors in the actual linguistic data is always possible even if they are controlled for. If such errors do exist, they are probably very few and would hence not cause any implications for the study as a whole. Translations of concepts do not always completely correspond to each other, brought up by Saeed (2003), Willners (2001) in sections 2.1 and 2.2.3, and insufficient knowledge of the languages used could cause difficulties. Furthermore, the cultural aspect of what a concept denotes cannot be completely neutral in any language cf. Saeed (2003) and Swadesh (1971). However, due to the very fundamental concepts used in this thesis, the potential problems rising from this ought to be limited.

Even if many of the examples featured in (Willners 2001), point towards binary oppositions as being fundamental for language there are some inherent problems; oppositional word pairs do not always fit the binary mould, even though binary components may be present within the

mould, e.g. 'sharp' could be related to both part of the binary pair SMOOTH and ROUGH creating a tripartite opposition in texture. The same situation applies to deictic terms; the most common distant deictic distinction beside the two-way is three-way, which affect the concepts HERE, THERE, THIS, THAT in this study, and additionally there are more complex systems mentioned in Diessel (2005). Furthermore, in three-way deictic systems the actual distance can be connected to either person (speaker, listener or other) or relative distance from the speaker. In time-deixis 'now' could possibly be used together with 'then' (past) and 'then' (future), in person-deixis, the pair I-YOU could be grouped with a third person pronoun or a demonstrative and 'we' could be contrasted with 'ye' (inclusive) and 'ye' (exclusive). Then there are even more complex semantic groupings such as tastes and colors which are very fundamental to the human experience but difficult to handle in binary terms. As brought up in section 2.2.3, sometimes one concept can contrast with two others, without the other contrasting with each other, as in the case of 'old', 'new' and 'young'. Obviously, important connections between concepts are overlooked when only binary pairs are used. Though if examples such as these above were to be included in a study like this new problems would arise, hence a new framework would have to be constructed in order to both capture the more complex semantic groupings and the precise binary connections.

The normal phoneme distribution from Wichmann et al. (2010) is based on 40 very common words, which due to most of them being short could produce erroneous phoneme distributions, though the featured concepts in this study also belong to the same type, and the 40 words were not chosen for potential phonosemantic form, but for their stability over time. Furthermore, the analysis of deviation from normal phoneme distribution could also be misleading in cases of sound groups having very low occurrences, particular in the Combination-grouping, as well as Voiceless Palatal and ə-like. Their ASJP transcription system was converted into the system used in this thesis, making small differences in sound classification to become larger errors (in percentages) when very small groups were added together, though these problems are marginal for the sound groupings and biplots. Also, using thresholds of 50/50 % and 100/75 % over- and underrepresentation always causes some data to be disregarded, despite almost reaching the threshold-level e.g. there were several cases of 45 % over- or underrepresentation, potentially interesting.

Consonant clusters were not taken into consideration due to difficulties in classifications of the different combinations of sounds. This might however have led to interesting connections

between certain concepts and clusters being neglected. The sonority-analysis would probably have been greatly influenced if clusters were included, since clusters of consonants often leads to different kinds of assimilations, affecting how the sounds are perceived, and the same would apply to diphthongs which were, just as the consonant cluster, segmented into phonemes in the study. Vowels could also be divided into several other way, e.g. more strictly according to sonority as in Hogg & McCully (1987:33).

When it comes to linguistic data, some linguistic forms used in the study were lacking, possibly making the total data less reliable. Holman et al. (2008) experienced similar problems with their data collection, resulting in a threshold of 70 % of the total numbers of concepts featured, and if this is taken to be a reliable figure, the data of this papers would be reliable in all cases. Even if the desired number of linguistic forms for all languages is 100 %, the timeframe for this paper did not allow a longer data collection period, and furthermore, in most cases the linguistic forms lacking were less than five per language, spread out among the language families and the different concepts.

5.2 Form and Meaning

The results showed systematic clustering of semantically related concepts purely based on sound correspondences. The majority of the investigated words contained phoneme distributions which deviated from the average, which means that the actual linguistic forms of the featured languages have to share similarities. This section presents factors which might be responsible for these associations between different meanings, as well as between sound and meaning.

5.2.1 Embodiment and Phonosemantics

Max Müller (1861) listed and criticized various onomatopoetic explanations of how language may have originated which flourished during and after the Age of Enlightenment, the so called "bow-wow theory" etc. However, the increasing evidence for phonosemantics playing an important role in human language, explained in section 2.3, together with the very closely related notion of embodiment, the shaping of the human mind by the human body, suggests that both of these act as strategies of concretization and grounding of more abstract and/or unknown concepts. Unsurprisingly given mentioned connection, gestures have been proposed as a steppingstone between simple vocalizations and full language. According to Premack & Premack (1983) gestural and vocal language are built upon similar neural systems, correlating with nonhuman primates which can be taught at least primitive communication in the form of

gestures or symbols, again illustrating the use of something concrete (physical gestures) for abstract concepts (spoken language). Bickerton (1990) proposed an early protolanguage with very little syntax emerging with Homo erectus around one million years ago. Pragmatic principles would govern clausal organization, then semantic principles, followed by syntactic principles, all without recursion. The lexicon would consist of lexical categories only, without functional categories and complex morphology, there would be little diversity in categories and a poor vocabulary. Bickerton refers to primates which in captivity are able to acquire language to some extent, though this only involve content words, cf. grammaticalization. Building on this, Muysken (2009) writes that the model proposed by Heine & Kuteva (2007) suggests gradual enfolding of the lexical system, which would point at the syntax and the lexicon would have evolve intertwined, while functional categories emerged at the interface between them. What can be deduced from this is that content words rather than functional categories were the keystones of language; more complex and abstract concepts are ultimately built upon words with real world referents.

Through neuro-, electrophysiological and metabolic imaging studies as well as ERP responses Pulvermüller (1999) showed that content words, function words, and words referring to actions and perceptions, have different neurobiological counterparts, i.e. sensory projections vary with semantic word properties. Neuroanatomical evidence from monkeys points towards the perisylvian cortex having long-range connections between areas anterior to motor, adjacent to primary auditory, and posterior to primary somatosensory cortex, which on a psychological level could be connected to embodiment of the phonological form of the words acquired during language acquisition. Concrete content words e.g. nouns, adjectives, and verbs have a strong connection to concrete as well as imaginable meanings, probably through these long-range connections. Function words on the other hand, e.g. auxiliary verbs, articles etc., have primarily a grammatical purpose which means that their meanings are not directly connected to objects or actions. According to Pulvermüller there is a continuum of meaning complexity in how well connected the mentioned different areas are between concrete content words with clearly defined referents, more abstract words which may or may not be used to refer to objects and actions and function words without any physical object referent. Furthermore there are words not referring to objects and actions in the same sense as e.g. “house” refers to an object. 'Joy' and 'anger' are connected to patterns of muscle activity, at least indirectly, meaning that that the degree of abstractness of an item is not the only relevant factor. Action words probably refer to the movements of one’s own body and thus used

frequently when these kinds of actions are being performed, e.g. when a meaning such as 'leg' is heard, the area associated with running and other types of movements are activated, hence the linking of word form in the perisylvian cortex to areas related to motor programs. Though, not all action-related associations involve the motor modality, e.g. 'to fly' or the 'the plane' are never connected to performances of the subject's own body but perceived visually. Perception words such as odors, tastes, sounds, visual perceptions and so forth would connect the perisylvian cortex to visual cortices, as well as temporal, and/or occipital lobes. Agreeing with this is that phonosemantics is sometimes said to be explained by synesthesia, i.e. neurological connections between the word's sounds and meaning, when the meaning is connected to other senses such as 'sight' or concepts perceived with several senses such as 'form'. Cytowic (1989) writes that synesthesia is an idiosyncratic phenomenon, grounded in a person's personal experience, arguing it is a more intense form of metaphoric speech. People perceiving synesthesic relationships between e.g. days of the week and colors or the taste of numbers might not be in majority, however synesthesia-related phenomena seem to be very common to the human experience. *Senseanalogies*, sense analogies that are linguistic metaphors, e.g. *dark tones, warm colors* etc., are rather common (Abelin 1999), and phonosemantics seem to affect most people. There are also synesthesic finding more general in nature, e.g. correlations between vowels and colors (Jakobson & Waugh 1979). Through electric stimulation Cytowic (1989) found a connection to the limbic system during subjects' synesthesic experiences, arguing that in some individuals the limbic system sometimes overrides the cortex, which causes the boundaries between the senses then disappear. In experiments testing words such as *maluma* and *takete*, as in Ramachandran & Hubbard (2001), Oberman & Ramachandran (2008) found that individuals with autism spectrum disorders perform worse than neurotypical children in connecting phonemic structures with visual shapes. They suggest this to be caused by impairments in multisensory integration systems, yielding autistic individuals social, cognitive, communicative, and motor symptoms.

Lakoff's (1987) image schemas, i.e. mental patterns that structures understanding of experiences are often described as very abstract structures such as PATH and VERTICALITY. Zlatev (2005) therefore introduces the *mimetic* schemas which are more concrete and grounded in the body image, derived from the uniquely human capacity of *bodily mimesis* (Donald 1991; Zlatev, Persson & Gärdenfors 2005). *Bodily mimesis* is described (simplified) as involving cross-modal mappings, consciously controlled bodily motions, iconic or indexical correspondence between body (part) including its motion and

action, object or event, and that the subject intends for the act to stand for some action, object, or event directed at an addressee. However, if the act is fully conventional there no longer is any bodily mimesis. This means that as for preverbal representations e.g. the mimetic schema JUMP would precede image schemas such as VERTICALITY, crucial for aspects of language acquisition as well as language evolution.

Related to these schemas as well as embodiment are conceptual metaphors, the conceptual metaphor theory originating in Lakoff & Johnson (1980) states that the conceptual metaphor is not purely lexical, it is a deep conceptual phenomenon shaping the way we think and speak. One of the most typical examples is "love is a journey", e.g. *Look how far we've come. We'll just have to go our separate ways. It has been a long and bumpy road.* Further examples include "more is up and less is down", "life is a journey", "social organizations are plants" etc. The actual metaphor is built up by the idea that a conceptual target domain is mapped onto a conceptual source domain. In the case of "love is a journey" the travelers constitute the source domain and the lovers the target domain, the journey is the evolution of the relationship and the obstacles encountered are the difficulties experienced, hence conceptual metaphors are grounded in experience and this affects language as well. The key here is a shared relative relationship in some sense, which also applies to phonosemantics. The observed usage of rounded sounds in the actual word for 'round' in many languages illustrates the shared form of the actual target concept and the vocal gesture imitating it, thus when the sound is produced whilst forming the vocal gesture a connection between the sound and meaning can be created.

A connection between sound and meaning seem to be functional in terms of memorization and learning. Kita, Kantartzis & Imai (2010) found that both Japanese and English children performed better in memorizing novel actions when phonosemantic conditions were introduced. Using concrete objects as substitute for an abstract concept in order to ease memorization has been used in rhetoric as well as in memorization competitions. Yates (1966:1) describes the technique of walking through an imaginary room or building a structure out of the contents of the speech as a way of memorizing long texts, used as far back as by ancient Greeks and Romans.

5.2.2 Oppositional Relationship

Several of the significant concepts of the study were found to be in different kinds of oppositional relationships, indicating that this is a important factor for phonosemantics and

maybe also to cognition. The names of spatial dimensions are in many languages based on the unmarked pole of oppositional adjectives; *length* from *long* and *width* from *wide* etc. Also, in questions the unmarked pole can be used neutrally e.g. *how long is this table?*, while *how short is this table?* would carry a presumption. It has been debated whether children understand the unmarked pole earlier than the marked pole due e.g. to higher frequency in usage, though the same ratio of usage is found in adults as well. Also if adults are asked to make a comparison between objects using the marked pole, it takes longer than if the positive is used. For more arguments, both for and against, see de Villiers & de Villiers (1978:139-141). Based on motion predicates acquired by the child during the "vocabulary explosion", from 16 to 24 months classified by Tomasello (1992), Zlatev (2005) found further evidence for the role of oppositions in that non-oppositional predicates corresponded to so called *mimetic schemas*, being dynamic representations of everyday actions and events, while the oppositional ones corresponded to "image schemas", indicating that semantic oppositions of language might be a prerequisite for some concepts, including basic ones. Regarding morphology, Cinque (2013) writes that while it is common to grammatically encode proximity (e.g. proximal and distal), size (e.g. diminutive and augmentative), number (e.g. singular and plural) etc., while apparently, no languages encode distinctions such as "strong/weak", "favorable/unfavorable" etc., and furthermore, these encodings usually come in oppositional pairs. Regardless if understanding of the unmarked poles comes before understanding of the marked ones or if they are acquired at about the same time, it seems evident that in order to reach a deeper understanding of contrasts, as well as language as a whole, opposites have to be acquired. Using contrasts in a semantic domain probably helps in understanding the actual frames of the domain, which would lead to quicker comprehension, which in turn leads to correct usage of each concept. It might actually be the case that the domain do not exist at all without having the oppositional pairs in place.

Hamlin, Wynn & Bloom (2007) found in several experiments that 6- and 10-month-old infants preferred individuals who help another to those who hinder another and neutral individuals, and they also preferred neutral individuals to hindering individuals. Preverbal infants being able to distinguish between individual's actions towards others could according to Hamlin et al. serve as the foundation for moral thought and action, which would make social evaluation a biological adaptation. Categorization of individuals seems to be closely related to oppositional thinking, which could also be responsible for the formation of social groups. Even though the world is far too complex to be described purely in binary terms,

perhaps pairs can be used in order to create larger networks which at least comes a bit closer to the actual state of affairs. This would in many ways simplify the various relationships of the world, but also make the world more understandable. A fitting example of this type of networks is Reay's (1994) network of phonesthemes, see Figure 11, illustrating how different sounds or clusters of sounds are fused together to constitute concepts, i.e. such as *sc-* 'light movement' and *-ump* 'rounded object or collection of objects'. The components can then be replaced, constituting related concepts, making their numbers grow larger over time and might be extended indefinitely.

Figure 10: Reay's (1994:4065) phonestheme network of English.

5.2.3 Semantic Origins

Based on neurological and memorization experiments, metaphors and so forth, it seems to be that meanings grounded in the body (both its appearance and abilities, but also unconscious processes) and in the surrounding world are the more fundamental type for both cognition and language. It opens up the possibility for certain, perhaps more general, meanings to presuppose other more specific meanings. From the more basic meaning, new meanings can be created grounded in various ways, e.g. *light* in English can refer both to luminosity, extended to color, and to weight, extended to movement (light, swift) and consistence (light, thin). According to Shimotori (2013:43-59) we do not perceive the world exactly as it is, since our senses work as biological constraints, filtering the input, e.g. humans cannot see

ultraviolet light while bees can, hence perceiving flower coloring completely differently. Experiences of what we already know has made us automatically and unconsciously code and categorize new input of perceptual objects and events, leading to (sometimes erroneous) associations and connotations. Furthermore, how we express information by using language does not exactly correspond to what we perceive either; it reflects how we understand the input information, e.g. the division of color terms differing among languages mentioned in section 2.2.

Adjective classes vary greatly among languages, some have very limited classes which contain very few members e.g. the North Australian language Malak Malak which has only seven words belonging to the adjective class (large, small, short, young, old, good, bad), while other have open classes containing hundreds, as many Indo-European languages, described in Dixon (1982:1-62). Dixon writes that all lexical items could fall into a number of (possibly universal) semantic types based on syntactic and morphological properties. Dixon lists seven semantic types for adjectives which seem to follow a specific order within the NP though without specific ordering within each type, obvious when *a fast, black, car is* contrasted with *a black, fast, car* the former sounding more correct than the latter. These semantic types also seem to hold on a universal level based on a study of 17 languages with small closed adjective classes. Adjectives types fall into three sets, the smallest adjective classes contain four semantic types, DIMENSION e.g. small, wide; AGE e.g. young, old; VALUE e.g. bad, proper; COLOR e.g. black, dark (Dixon 2010:73-76). Medium size classes contain an additional three types, PHYSICAL PROPERTY e.g. hard, heavy, hot; HUMAN PROPENSITY e.g. jealous, happy; SPEED e.g. quick, slow, and the largest classes contain a total of 13 types, also featuring DIFFICULTY e.g. difficult, simple; SIMILARITY e.g. similar, other; QUALIFICATION e.g. true, usual; QUANTIFICATION e.g. all, few; POSITION e.g. high, distant; CARDINAL NUMBERS including first, ordinal numbers. Small adjective classes have almost all of their members from the four core types e.g. Igbo, though Dixon further writes that specificity of the semantic types varies among languages; the Sango word *kótá* corresponds to English *big, wide,* and *thick*, while *kété* corresponds to *small, narrow* and *thin*, corresponding at least partially to the Large and Organic, and Small and Intense Vision-Touch clusters. Considering the fact that the four core types seem universal, it does not seem too farfetched to assume that there are fundamental oppositional pairs more basic than others, in some ways similar to Wierzbicka and Goddard's *Semantic Primes*. Wiendold & Rohmer (1997) showed through a typological study that the priority of

lexicalization of dimensional expressions occurs in a fixed order, from the more general meaning to the more specific meaning, composed by two interlocking scales originating from SIZE, Figure 12.

SIZE → LENGTH → DISTANCE → DEPTH → HEIGHT
↓
THICKNESS
↓
WIDTH

Figure 11: Lexicalization of dimensional expressions based on Wiendold & Rohmer (1997).

Clark's (1973) Semantic Feature Hypothesis assumes that when children begin to use identifiable words they do not know the full meaning of them, instead they know only some of the features or components of meaning compared to adults' lexicon which means that children start off by using words with fewer features i.e. more simple, general words such as 'big'. Barlett (1976) examined the acquisition of *big*, *little*, *tall*, *long*, *short*, *wide* and *narrow* in English concluding that more general terms describing overall size are acquired prior to those which describe length and width. de Villiers & de Villiers (1978:121-150) write that the first words understood by children are proper names, which have only one referent for each word such as *mommy*, *daddy*, favorite toy's names etc. These are followed by common nouns, which are more complex in the sense that they refer to a whole class of objects. The same difficulty is found in learning simple verbs and adjectives since the properties they refer to can be shared among many objects. The next step is relational words such as dimensional adjectives which depend on some kind of standard of the referent, i.e. a big ant is not as large as a big elephant. Deictic expressions become even more difficult since they also take in the speaker as a referent when talking about time and space. Furthermore, children generally overextend words' meanings before fully comprehending all their semantic features, and in different ways, e.g. *dog* can be used not only for a pet and other dogs, but also for horses, cows, cats, etc. if having four legs is in focus, while if the furriness is the most salient feature for the child *dog* might be used for things such as wooly blankets. de Villiers & de Villiers write that children create semantic categories based on similarities in perceptual or functional attributes e.g. it has been observed that *heavy* is used for any physical exertion, actually not involving weight. The same applies to priming effects, i.e. that it is easier to think of e.g. 'wolves' if the previous stimuli was 'dogs' than if it was 'grass' due to the canines belonging to the same semantic field, while the similarities between 'wolves' and 'grass' are very few. Also, in language acquisition 'blue' is rather easy to learn if another color already has been acquired

since the frames in which colors function is already known, it is only the type which new. Based on various experimental results presented in de Villiers & de Villiers (1978) BIG and SMALL are the most frequent and easily learnt dimensional adjectives since they can be used along any or all three dimensions, i.e. requiring only general physical extent, while TALL-SHORT and LONG-SHORT require vertical or horizontal dimensions and WIDE-NARROW and THICK-THIN require further specifications still. Furthermore, children often give *little* or *small* as opposites of *tall, long, wide* and *thick,* and *large* or *big* as opposites of *short*, *narrow* and *thin,* but seldom the other way around. These dimensional adjectives are hierarchally ordered acording to how difficult they are to acquire based on how frequently they are used by both adults and children and semantic complexity primarily based on Clark (1973), Figure 13, aligning rather similarly with the findings of Wiendold & Rohmer.

BIG/SMALL → TALL/SHORT, → LONG/SHORT → HIGH/LOW → WIDE/NARROW, → THICK/THIN → DEEP/SHALLOW

Figure 12: The hierarchy of dimensional adjective ordered by difficulty of acquisition based on Clark (1973).

Similarly, Viberg (1984), in Nakagawa (2012), found a potentially universal perception verb order as well based on 50 languages. Viberg proposed that semantic extensions across modalities are unidirectional, going from higher to lower modalities, i.e. from unmarked modalities to relatively marked modalities; the most basic meaning being 'see', followed by 'hear', 'feel', 'taste', and 'smell', the latter always implying the former, Figure 14, though steps can be skipped, 'see' can be extended directly to 'feel' without going through 'hear'. The hierarchy was revised in Viberg (2001) placing 'feel', 'taste', and 'smell' on the same level, though still keeping the implicational order of 'see' followed by 'hear' and the last three.

SEE → HEAR→ FEEL, TASTE, SMELL

Figure 13: Perception verbs ordered by lexicalization, based on Viberg (1984).

When looking at the semantic typology of ideophones, it becomes evident that they cover broad semantic areas and a wide range of sensory imagery such as sound and movement, as well as visual patterns, shapes, tastes, textures, inner feelings etc. (Dingemanse 2012). Dingemanse further suggests an implicational hierarchy for ideophones as well, Figure 15,

starting off with sound, i.e. onomatopoeia, as the most fundamental and through other sensory modalities ending with inner feelings and psychological states.

SOUND → MOVEMENT → VISUAL PATTERNS → OTHER SENSORY PERCEPTIONS → INNER FEELINGS AND COGNITIVE STATES

Figure 14: The implicational hierarchy of different types of ideophones based on Dingemanse (2012).

The hierarchy is constructed by an interplay of multiple factors; sensory systems, the sensory input from the environment, and semiotics of depicting sensory imagery in speech. SOUND is both common and salient for humans, and can be depicted by imagic iconicity, the simplest kind of semiotic mapping according to Dingemanse. Furthermore MOVEMENT often comes together with sound in sensory input, placing it on the second most fundamental level. VISUAL PATTERNS e.g. spatial configuration and surface appearance are also common observable and relatively salient and share suprasensory attributes with speech, which is also true for OTHER SENSORY PERCEPTIONS, although these are probably less directly observable. This last state could then function as a bridge to extend to INNER FEELINGS AND COGNITIVE STATES which are the least directly observable but shares some of the same suprasensory attributes.

5.3 Explanatory Suggestions for the Semantic Relations

It is not our modern society which has given our associations between various concepts, it is the surrounding world in which we have evolved. The need for describing and understanding the world has forced us to try to classify the elements of nature. We however are not capable of accurate physical descriptions since all of our input is filtered through our cross-modal sensory perception, existing knowledge gained from life experience, culture as well as what is imprinted on us as infants. We perceive that it is the sun that is moving across the sky, not the Earth moving around the sun since it seems logical based on what we know about the world and it is in this context language has evolved. As described in the results, the semantic domains found do have some members occurring in places other than the expected in some of the biplots, though the overall division clearly shows definite semantic clusters, hence not too much can be said about the sound groupings' respective effects of capturing different types of concepts. In some cases clustering might only be the result of phonological coincident, though the actual linguistic forms of these concepts are still significant, seeing as they are located outside of the center of the biplots. Also, the same phonological resource seem to be used to denote different meanings, but are distinguishable when the phonological makeup of their

oppositional concepts are taken into account. I here propose explanatory suggestions for the connections between the concepts constituted solely by phonetic makeup.

Parent:

Out of the concepts occurring outside of the center MOTHER and FATHER were always located furthest away, not only from the center but also from each other, yielding the clearest distinction between all of the significant concepts in the study. I, or EGO, was also situated close by MOTHER, further discussed in 5.4.1.1.

Small & Intense Vision-Touch:

SMALL and SHORT are very closely related semantically, basically denoting the same thing with the difference of SMALL being more general and SHORT concerning vertical and horizontal smallness. The traits of these concepts can easily be mapped over to the distance-dimension, i.e. 'small distance', hence NARROW and NEAR, and as well on to the quantity-dimension, i.e. 'small quantity', including FEW.

BLACK, WHITE, DARK, HOT, COLD and WET includes the domains of color, temperature, light and dampness and in the case of color and temperature both poles of the domain, i.e. both concepts are present, BLACK-WHITE and HOT-COLD. Color, the term being a bit misleading in this case, and light are quite obviously connected; darkness is more black and light is more white. However LIGHT (not DARK) which should be paired together with WHITE is lacking in this cluster though occurring in the same region as the Large and Organic concepts. Temperature could probably be connected to the light-domain though heat generated by the sun and fire, connecting heat (HOT) or lack of heat (COLD) together with WHITE, BLACK and DARK. The dampness concept, WET, might be connected to lack of light and lack of heat in a similar manner. And the tactile properties of ROUGH and HARD probably connects them to the rest of the Intense Vision-Touch concepts. The connection between ROUGH and HARD and the Small concepts might be explained through stones, pebbles and pieces of wood. Occurring everywhere in nature and used by us and our ancestors as tools for millions of years these materials have been a very important element of the surrounding world. Those used are often small compared to many other things of note and use in nature such as game, hills, bushes etc. since in order to be used the stones need to be able to be lifted in one hand, possibly except for some larger ones used for crushing which would still be able to be lifted by two hands. Even larger kinds of stones might be seen as integrate,

immovable parts of nature, together with trees, mountains etc. This way of reasoning relates to the various studies and writings concerning *Ecological Psychology* by James J. Gibson, who argues that perception is based on what information, gathered from experience, we have gained from the world (Gibson 1977; Mace 1977). Applying to both humans and animals, the *life world* powerfully affects the behavior, what action are possible and how the world is perceived, connecting the organism to the environment.

Except for the possibility of the rest of the Intense Vision-Touch concepts simply utilizing the same kinds of sounds as the Small concepts, they might, similar to OLD, be connected via ROUGH and HARD i.e. via the perception of texture. Whorf (1956) writes about similar associations, between bright, cold, sharp, hard, high, light (in weight), quick, high-pitched, narrow and so forth, as well as associations between dark, warm, yielding, soft, blunt, low, heavy, slow, low-pitched, wide, etc. And furthermore, sounds cried out when experiencing small quantities of intense pain such as placing a hand on a hot stove, the feeling of a cold hand on the back or being blinded by bright light are usually high pitched, which could be a indexically associated between sound (which are also used for the Small concepts) and meaning in this case. BLACK and DARK would then be included by association with the other Intense Vision-Touch concepts. Another explanation for the similarities in phoneme distribution shown in the biplots could be that these concepts are connected to danger, which according to Ohala's frequency code would fall in line with acting small and submissive. This also correlates with one of the most salient examples of deviating phoneme distribution, the overrepresentations of the Voiceless Alveolar sound group in these concepts which are high-pitched. Though this is far from the only sound group which plays a role for the placements in the biplots, hence the Deviation from Average results, Table 15, should be consulted every time the concept, clusters and their relations in biplots are considered.

The Small concepts could be related to OLD through ROUGH and the other Intense Vision-Touch concepts, since older things usually become more rugged, less smooth and more imperfect. EMPTY alone could be associated with smallness, i.e. 'small quantity', though it might as well be connected to largeness if it is perceived as a large void, and the same situation in reverse could be applied to FULL. Regardless of which, both Containment concepts occurring rather close to the Small concepts etc. points towards a purely coincidental, though not trivial phoneme distributions in these concepts.

Large-Organic:

BIG, LONG and DEEP are related in the same manner as their counterparts SMALL and SHORT i.e. BIG is the more general concept, while LONG and DEEP constitute the vertical and horizontal versions of the same concept. The very close connection to the organic concepts ROUND, SOFT, SMOOTH and THICK found in the biplots could be explained again through the surrounding world, i.e. large things of note are usually either hills, mountains and the like, or animals which usually do have the traits of the organic concepts. Furthermore large things at a distance are usually perceived as having smoother shapes than they really have. A craggy mountain ridge can be mistaken for a cloud at the horizon if the distance is great enough. This proposition also helps to reinforce Small and Intense Vision-Touch connection since things examined at a very small distance allows imperfections to be seen leading to small things, which are not seen at a large distance at all since they are lost in crowd of larger elements of nature, to be perceived as more rugged and rough. Though it has to be noted that this is mere speculation. Also, the most salient overrepresentations of these concept were the Trill/Tap/Flap and Lateral sound groups and perhaps the Voiced Velar group.

Horizontal-Vertical Distance:

FLAT, SHALLOW and WIDE/BROAD all denote a shape which is extended in one dimension, a surface and very thin in the other dimension, the primary overrepresentation was the Lateral sound group. Besides from the obvious explanation of simply being the counterpart of the same shape as the first three concepts, though in a vertical manner the vertical concept ABOVE could fit into this domain by being seen as something being above, over, on top of something else which is reliant on the entity below in order to be defined. This could then be seen as something running along something else, yielding the connection to the extended dimension, cf. Lakoff's (1987) image schemas. Though it should be noted that ABOVE occurred slightly farther away from the rest of the members of this cluster, possibly using WIDE/BROAD as a bridge, and might therefore not be as closely connected semantically either. THERE is the only significant deictic concept besides I (EGO) which is not found within the Deictic cluster in the biplots. The term can be used in both horizontal and vertical concepts and is at a distance, giving the connection to the concepts above. Furthermore THERE being deictic involves following a line towards the referent, quite similar to the extended dimension in the other concepts.

There could be a connection between the Horizontal-Vertical Distance concepts and SLOW in the sense of something being stretched out far away, hence taking a lot of time to reach or to construct. This explanation would however semantically connect SLOW to the Large-Organic concepts which is not the case, indicating a coincidental likeness in phonological makeup of the different concepts. The connection to LIGHT (not HEAVY) and the Gender concepts are probably coincidental as well.

Deictic:

All deictic concepts except for the already mentioned THERE and I which is located far away from the center occur in the same area, and seem to have overrepresenations in the Voiced Palatal sound group and underrepresenations in the Voiceless Labial sound group. YOU and THAT are logically connected, both being non-first person and also away from the speaker in terms of distance. THIS and HERE could be associated with the speaker, though they are just as YOU and THAT indicatory of something which is not the actual speaker. The indicatory deictic concepts, occurring together with the Large-Organic concepts, can be related through the notion of 'large distance' i.e. using "large sounds" for denoting something that is not the ego.

Containment & Gender:

The Containment-domain, FULL and EMPTY and the Gender-domain, WOMAN and MAN, have their concepts being located very closely together despite being two poles of the same domain, further discussed below.

Diurnal:

DAY and LIGHT (not DARK) of the Diurnal-domain occurring together is probably due to common lexical origin in many languages, however these were located outside of the center in the biplots as well, pointing towards unusual sound-makeup. Their quite close co-occurrence with the Large-Organic and Deictic concepts is rather unclear, probably just a phonological coincident.

OLD:

OLD had obvious semantic relation with other concepts, except for possibly Small and Intense Vision-Touch, seeing as it also is overrepresented in the Trill/Tap/Flap sound group. Interestingly OLD was also present among the 23 stable words in Pagel et al. (2013), while

YOUNG was not, the same as in this present study. One explanation could be that OLD is the oppositional pole of both YOUNG and NEW.

Non-significant Concepts and Relations:

When comparing the significant concepts to the non-significant concepts some traces of patterns can perhaps be seen. GOOD-BAD and BEFORE-AFTER are very abstract, maybe even too abstract, which could prevent using the surrounding world as reference in order to make someone understand a concept, e.g. pointing towards the sky for a notion such as 'up'. This does however only account for four concepts, and HIGH-LOW not being significant argues against this suggestion. One could then suggest that dimensional concepts seem more fit for phonosemantics based on the results, which seems to be true for the most part, though once again HIGH-LOW is an exception. The author cannot either explain why MUCH/MANY, FAR, NIGHT, NEW, DRY, THIN, HEAVY and FAST are not significant while their antipodes are.

5.4 Oppositional Relations between Clusters

5.4.1 Separated Components

5.4.1.1 MOTHER-EGO & FATHER

The most clearly defined concepts in terms of oppositional pairs were MOTHER and FATHER. Both are located far away from the center but also far away from each other, which makes the fact of them being so closely semantically related very interesting. Besides the simple syllable structure in which the vowels are usually [a] of some variety, CaCa, MOTHER and FATHER both use types of stop sounds for their consonants; nasals and plosives. There are however very important differences; MOTHER is associated with voiced continuous sonorants while FATHER is associated with voiceless abrupt obstruents. According to Swadesh (1971:191-199), aside from 'mother', nasals are common in words for 'aunt', 'grandmother', 'old person', 'baby' etc., as well as for 'female 'breast', 'teat', 'nipple'. The association might seem obvious in a purely semantic sense, though phonologically speaking babies usually produce nasal sound while breastfeeding. While the actual action not require any sounds at all, if the baby would vocalize whilst breastfeeding, the double closure of the mouth, lips and tongue in the back, forces sounds to become nasal, also reflecting contentment-sounds often expressed when eating, e.g. [m:], and furthermore the most relaxed form of human vocalization is nasal. Adding to this, the Niantic form for BREAST in Wichmann et al. (2010) resulted in *muma*, in which the first three sounds are articulated with

the lips suggesting that it reflects the suckling of a child. The connection between 'breast(feeding)' and 'mother' could then be extended to other female relatives. Also, Wescott (1971) suggests that the [m] in 'murmur', 'mouth' and 'milk' may be connected in a similar way, and going outside Indo-European, Estonian *mokk* 'lip', *maitse* 'taste' and *musi* 'kiss', and Basque *musu* 'kiss' follow the same pattern.

Swadesh also suggested that based on physical characteristics of the sexes, the mother would be seen as the *soft* parent and the father as the *tough* parent. This is a rather bold statement, though there might be a grain of truth if attention is once again directed back to breastfeeding (see also Hrdy (2009) for cultural perspectives of fathers not being the primary care providers of infants when compared with mothers and even other relatives in some cases). During the feeding the infant might suggest the mother with more than safety and relaxedness (= relaxed velum), the sensation of 'soft', 'smooth' and 'round' and maybe even 'large' (the mother certainly being larger than the infant) could also be extracted. Despite these concepts occurring much closer to the center than MOTHER in the results they are all located in the same general area, i.e. towards the same direction in the biplots, compared with e.g. the Small, Intense Vision-Touch and Horizontal-Vertical Distance clusters suggesting at least partially over- and underrepresentation of the same phonemes. The father, on the other hand, who does not breastfeed does not get associated with these sensations and could then, perhaps, be connected to the reversed - tensed sounds and feelings. The close connection between the ego and the mother through breastfeeding during the early stages of postnatal life, optimally occurring for at least six months (Kramer & Kakuma 2009), could perhaps make the two concepts intertwined enough for them being perceived as the same concept by the infant.

5.4.1.2 EGO & Deictic

The first person singular personal pronouns cross-linguistically contains nasals as shown by Traumüller (1994), also suggested by Swadesh (1971:199) and confirmed in the results of this paper as well. Even in Indo-European languages which usually employ [k] or [g] for this purpose as a common sound, the majority of the oblique forms contains nasals as well. Out of the 75 featured languages 50 contained at least one nasal in first person singular personal pronouns and another six languages had at least one nasal in their oblique forms; Cebuano, Tetum, Thao, Kavalan, Seediq and Vietnamese. Nyanja, Nama and Zinacantán Tzotzil had no data of first person singular personal pronoun at all, and seven languages, Ghulfan, Kanuri,

Ayoreo, Yaminahua, Aché, Wapishana and Catuquina, had no data of the oblique forms. Nine confirmed languages had no nasal in any form; Hawaiian, Tongan, Malagasy, Malay, Mandarin[5], Tlingit, White Hmong, Ket and Japanese. Disregarding the three languages with no data, and counting the seven languages with no data of oblique forms together with the nine languages without any nasal (despite the possibilities of nasals in the oblique forms), this leaves 56 languages with nasals in these concepts (77,8 %) and 16 without nasals (22,2 %). Traumüller suggested two possible explanations this association; the lack of lip protrusion of e.g. nasals would avoid pointing towards something, while plosives and rounded sounds would indicate non-EGO deictic concepts. Or that the "first person-sounds" would be connected to relaxedness and safety, i.e. one is sure of one's own intentions, while plosives and the like would be associated with tenseness, i.e. insecurity of others (cf. Ohala's frequency code), aligning with the MOTHER-FATHER relation explained above. The explanation for nasals occurring only in the oblique form could be that children might use it more frequently than other form since children might refer to themselves as recipients of actions.

Unsurprisingly the EGO is acutely separated from the concept of the Deictic cluster. In deictic systems, regardless of type of system, the one point of reference which is always included is naturally the speaker. Using oneself as a frame of reference is probably imperative in order to understand others and others' frames of references. Having a so called *Theory of Mind*, the ability to represent, conceptualize, and reason about mental states, yields the ability to distinguish between how oneself thinks and how others think (Malle 2002). This insight makes it possible to adapt to being a part of a group, understanding social codes and interpreting other's feelings, as well as reacting in a suiting manner towards them. Not until the second year of life are children able to understand that others than themselves experience psychological states, which can differ from their own. They then also become aware of themselves and recognize themselves in a mirror (Brownell, Zerwas & Ramani 2007). This self-awareness revolution giving the crucial cooperative skills that are needed to function socially paradoxically leads to the infants being more and more autonomous, independent and detaching themselves from their mothers (Moore 2006). This suggests that since the world of an infant is very egocentric and understanding of others is limited for quite a portion of early life, a division (both phonosemantic and purely conceptual) between EGO and all other referential concepts, including the THIS, THAT, YOU, HERE etc., is rather natural. Two-

[5] In Mandarin the inherent nasal [ŋ-] has been lost, cf. Cantonese [ŋɔːɻ].

year-olds understand the switch in perspective concerning the deictic expression here and there, although it is unclear when full comprehension is acquired (de Villiers & de Villiers 1978:144-145). The contrast *my*/*your* was rather easy for the children to understand since it did not involve any distance contrast, simply a switch from speaker to hearer. Autistic children on the other hand had great difficulties in handling these basic distinctions, instead referring to themselves as you echoing the speakers reference.

Heading back to what is actually visible in language, what is encoded in pronominal systems differs among languages more than what one might think, not only regarding gender and number, but also how references to speech participants are treated. The Wari' language, spoken in the state of Rondônia, Brazil, does not actually contain any pronouns in the traditional sense (Everett 2005a). Instead it has a system of spatial and temporal demonstratives, making the pronouns periphrastic. They are constructed by combining a proclitic e.g. *co* 'singular masculine' with a demonstrative e.g. *cwain* 'distal', creating *co cwain* 'that: distal, singular masculine', which would correspond to 'he', Table 16. For third person this might not seem too strange, however considering the fact that 'I' (masculine) is produced by saying *co cwa'*, 'masculine singular, proximate to speaker', this system is indeed very unusual typologically speaking.

Table 16: Paradigm of the spatial demonstrative pronouns in Wari', based on Everett 2005a.

	Proximate to Speaker	**Proximate to Hearer**	**Distal**
Masculine singular	co cwa'	co 'ma'	co cwain
Feminine singular	cam cwa'	cam 'ma'	cam cwain
Neuter	'i ca'	'i 'ma'	'i cain
Plural	caram cwa'	caram 'ma'	caram cwain

While Wari is lacking formal first and second person pronouns, functionally, the demonstratives denoting proximity to speaker and hearer could be mapped onto the first and second person concepts without any real stretch of the imagination, demonstrating how closely connected our perception related to speech participants and distance is. About one half of the sample of 234 languages found at WALS use two-way spatial deictic systems while over one third use three-way systems, which are either connected to speaker, hearer and other, or simple spatial gradation. Kemmerer (1999) showed evidence from both human neuropsychology and primate neurophysiology, suggesting that the visual system creates a basic distinction between the area around the body within a perimeter of roughly an arm's

reach, and the region outside of it. He concluded that the ways in which experience is structured for linguistic communication do not necessarily completely reflect how visual perception and motor control is structured. They are at least to some degree based on abstract semantic notions, which in turn seem to be grounded in visual perception and motor control. Perceptual representations of egocentric space seem to be related to motor control; peripersonal space is used to program movements of the arms, hands and head, while extrapersonal space is used to program visual examination. This could account for the indicatory Deictic concepts to be grouped together since HERE, YOU etc. are usually outside an arm's reach. Furthermore, in the Deictic cluster HERE is contrasted with THERE, found in the Horizontal-Vertical Distance cluster. This creates a very clear oppositional relationship between the two similar to that of EGO and the other Deictic concepts.

5.4.1.3 Small, Large & Horizontal-Vertical Distance

There is a three-way oppositional relationship between the Small and (possibly together with Intense Vision-Touch), Large-Organic and Horizontal-Vertical Distance clusters in the spatial dimension. Phonosemantically speaking, it is not hard to imagine associations between the three most distinctive vowels [i], [u], [a] and representative concepts for each of these clusters, SMALL, BIG, FLAT based on previous experimental results. These general kinds of shapes might constitute types of basic dimensions necessary for human thinking, i.e. make it possible to describe and understand most elements of the surrounding prehistoric world. Large-Organic could account for mountains (smooth looking when seen from a distance), valleys, clouds, larger animals (in particular mammals), the sun, the moon etc., while Small and Intense Vision-Touch could account for stones, pieces of wood, smaller animals (in particular insects), pieces of bone, stars etc. The Horizontal-Vertical Distance concepts could then be used to describe plains, oceans, lakes, the horizon etc. Combining the various semantic properties of these concepts clusters, though their phonosemantic associations (similar discussion found in Swadesh (1971: 206-211)), could yield small, round things i.e. berries, fruit, eggs, flowers; big, hard things i.e. cliffs, icebergs; small, flat things i.e. puddles, leaves and so on. Also supporting this suggestion is that ROUND is found in the Large-Organic cluster and FLAT in the Horizontal-Vertical Distance cluster. This oppositional pair could both be contrasted with 'pointy' or a similar quality which could align with the Small and Intense Vision-Touch clusters, illustrating the dimensional triangle.

According to Bowerman (1996) spatial categorization for visual or semantic similarities could be needed in order for linguistic counterparts to emerge, which is supported by various experimentally proven arguments. There is evidence for that prelinguistic children know a lot about space; they have highly constrained learning mechanisms which enable them to construct generalizations about objects (Needham & Baillargeon 1993). There is a close connection between linguistic and perceptual organization of space; perception and conceptualization of locations of objects are constrained by biology, e.g. top-bottom and front-back symmetry and their physical environment, e.g. working of gravity, hence children's spatial knowledge is reflected in semantics (Clark 1973). Various studies have also shown that spatial morphemes emerge after non-linguistic spatial knowledge is in place; children play games which require understanding of 'containment' and 'support' before acquiring the words *on* and *in* (Clark 1973). Furthermore it has been argued by Talmy (1983) and Landau & Jackendoff (1993) that closed class spatial morphemes encode only schematic information such as main axes and trajectories and not Euclidean information e.g. angle, distance or exact shape, though Brown & Levinson (1993) showed that closed class positional verbs in Tzeltal include a number of Euclidian properties of figures. This points towards the possibility of a small number of shapes being very basic for human cognition and the three-way oppositional relationship found in this study could constitute or at least be related to some of these very basic spatial categories.

5.4.1.4 Small-Intense Vision-Touch, & Large-Organic

Between the Small and Intense Vision-Touch, and Large-Organic clusters there are four oppositional pairs; two spatial and two textural, SHORT-LONG, SMALL-BIG, SOFT-HARD and SMOOTH-ROUGH, indicating a binary opposition between the concept clusters. SHORT-LONG and SMALL-BIG could be regarded as more or less the same concepts, hence they are probably perceived and treated in a similar manner as well. Though SMALL and BIG are possibly more basic in meaning, considering that SHORT and LONG are size concepts in the vertical or horizontal dimension. SOFT-HARD and SMOOTH-ROUGH also correspond to each other. In the natural world, SOFT and SMOOTH often co-occur; things that are soft are often smooth, and vice versa, and the same applies to HARD and ROUGH. Things that are both rough and soft are not really encountered, and even though smooth, hard things do exist, such as polished surfaces, these are often manmade, making the connection less intuitive.

5.4.1.5 DARK & LIGHT (not DARK)

DARK is unsurprisingly found in the Intense Vision-Touch cluster, while LIGHT (not DARK) (together with DAY) seems to be located close the Large-Organic and Deictic cluster, without any obvious connection to either of these domains. No clear explanations were found for the phonosemantic association between the oppositional concepts. Despite the fact that DARK is located together with other Intense Vision-Touch concepts such as BLACK which is contrasted with WHITE within the same cluster, it does not follow the same pattern.

5.4.2 Conjoined Components

There are two different ways of making a phonosemantic contrast between concepts of the same semantic domain. The first way is using phonemes in a word connected to a semantic domain, which seems to be the case for most significant concepts in this study, e.g. *maluma* for Large-Organic concepts and *takete* for Small and Intense Vision-Touch concepts, cf. Usnadze (1924), Köhler (1930), Ramachandran & Hubbard (2001) and Ahlner & Zlatev (2010). However some of the oppositional pairs were situated very closely together in the biplots, and still being positioned away from the non-significant center. This can probably be explained by simply sharing a common core and using a single phoneme as contrast (cf. the method of contrasting deictic demonstratives phonosemantically in Johansson & Carling submitted) instead of using the phoneme qualities of several segments of the word as in the lexical forms for SMALL and BIG etc. However, the underlying reason for this could also be that the pairs are not perceived as being two poles of the same domain, but two types of the same domain. The Gender concepts, being a cluster of its own includes both WOMAN and MAN and the same situation is true for the Containment concepts constituted by EMPTY and FULL. The co-occurrence of WOMAN and MAN can probably be credited to a shared lexical core being inflected for male and female, e.g. Hebrew אִישׁ [ʔiʃ] 'man' and אִשָּׁה [ʔiʃa] 'woman', though not necessarily in a grammatically active sense for all languages. A similar situation is probably true for the Deictic concept THIS and THAT, co-occurring in the Deictic cluster. In many languages only one sound constitutes the difference between the lexical forms of these concepts as in Nyanja [uju] and [ujo], Akha [hə] and [tə], Fataluku [e] and [i] and so forth, i.e. a shared core is used for both words, with a single phoneme used as contrast, making the words "inflected" phonosemantically. This yields a small, but perhaps relativity speaking, larger difference between the concepts, making them both identifiable as connected, i.e. sharing the stem, but also separated, i.e. denoting opposite poles of the same domain. A

similar case is MOTHER and FATHER, which with their possible shared CaCa-structure also could be viewed as inflected variations of the same stem[6].

FULL and EMPTY seem to be treated the same as WOMAN and MAN, though intuitively speaking, these oppositional concepts are more like§§§ poles of the same domain as opposed to WOMAN and MAN, with regard to semantics. Even if this is the case, the actual sounds used in the linguistic forms were different enough to be located outside of the center in the biplots i.e. having an average phoneme distribution. It is difficult to imagine that the shared root idea would apply to FULL and EMPTY as well, WOMAN and MAN are both varieties of humans, while FULL and EMPTY are diametrical states. And regarding HOT and COLD, the feeling of extreme heat and cold when touching something can initially be hard to keep apart, the feeling could simply be described as 'intense touch' or something similar. This is a possible explanation for their co-occurrence in the Intense Vision-Touch cluster. BLACK and WHITE, co-occurring in the same cluster, would have a similar explanation, i.e. that these color-poles could be hard to differentiate is perhaps more farfetched unless some kind of association with light and darkness is applied. Too bright light can certainly make one go blind, at least for a short period of time, similar to the perception of being a completely dark place. Furthermore, the Proto-Indo-European root *b^h*leg*- 'burn, shine' occurs in many words connected to light in the daughter languages e.g. Latin *fulgō* 'to lighten, glitter, shine' and Tocharian *pälk*- 'to shine, burn'. However in English the cognate of these words is *black*, the opposite of 'white' and 'light', which suggests that the concepts of 'fire' and 'burning' could be responsible for associations between LIGHT (not DARK) and DARK.

5.5 Linguistic Primitives

Based on the clearest semantic clusters of the biplots following the discussions above and taking the fact that the relative positions of the clusters varied very little among the different sound groupings into consideration, some suggestions for potential linguistic primitives can be made. Despite having a conservative perspective, certain semantic concepts and relations include MOTHER-FATHER, the three degrees of deictic distinctions and the tripartite shape-related concepts (linked together to the deictic distinctions via THERE in the Horizontal-Vertical Distance cluster, Figure 16. Large-Organic and Small and Intense Vision-Touch (which could include OLD) clusters are further backed up by the typological findings by Wiendold & Rohmer (1997) and the findings in language acquisition by Clark (1973), which

[6] The structure is also present in less apparent cases e.g. the English form *mother* and *father*; [mə-ðər] [fɑ-ðər].

both showed that SIZE is the foundation of other dimensional concepts. The deictic distinctions could be directly connected not only to first and second, but also to third person which was not featured in this study, reflecting subjective, intersubjective and unknown/other, cf. Table 16. There are also some relations which could be interpreted as concepts at least partially utilizing the same phonological resources but also being significantly deviant from the average phoneme distribution to possibly have primitive-like qualities, including; MOTHER and I (the EGO-part of the deictic distinctions), Small and Intense Vision-Touch, and lastly Deictic and Large-Organic. WOMAN-MAN (Gender-domain), FULL-EMPTY (Containment-domain) and DAY and LIGHT (not DARK) (Diurnal-domain). And the same applies to LIGHT (not HEAVY) without any apparent connections to anything and SLOW which could be connected to the Horizontal-Vertical Distance cluster although this should be considered a remote possibility. However, they do not seem to be related to the other semantic cluster in any obvious way. If there are shared qualities, they might be found if more concepts were featured in the study.

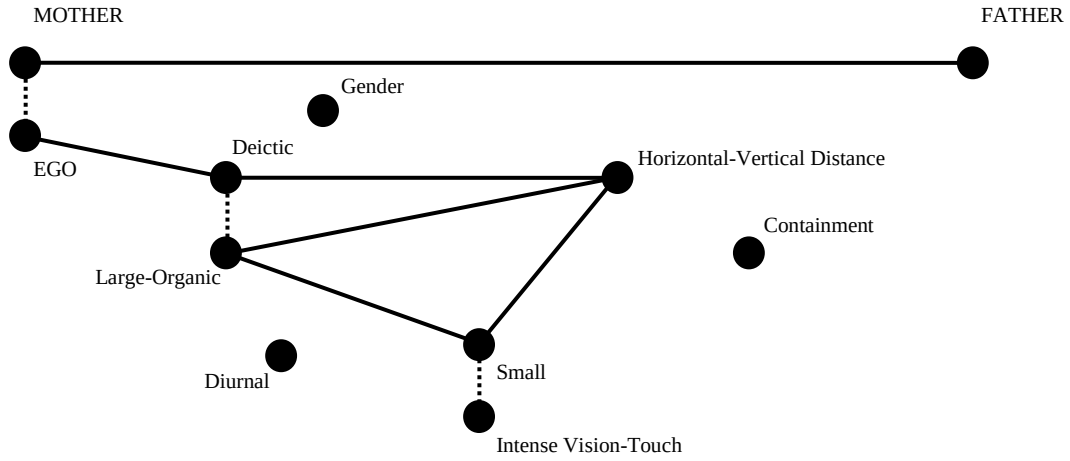

Figure 15: Conservative suggestions for potential linguistic primitives and relations. Full lines indicating certain relations and dashed lines indicating utilization of the same phonological resources between domains.

If a more generous and speculative approach is adopted, interpreting relations which might be utilizing the same phonological recourses as belonging to the same domains, a more simplified schema is found, Figure 17. MOTHER and EGO as one concept (via e.g. breastfeeding and feeling relaxed etc.) contrasted with FATHER in one direction, and constituting one of the extremes of the three deictic distinctions which are intertwined with two of the shape-related concepts; Deictic co-occurring with Large-Organic (via LARGE DISTANCE) and Horizontal-Vertical Distance including both THERE and spatial concepts. The two shape-related concepts further form a triangle, relating to Small and Intense Vision-

Touch if these are considered to be one (via e.g. protection oneself against fear, such as intense heat or similar, by seeming small etc.).

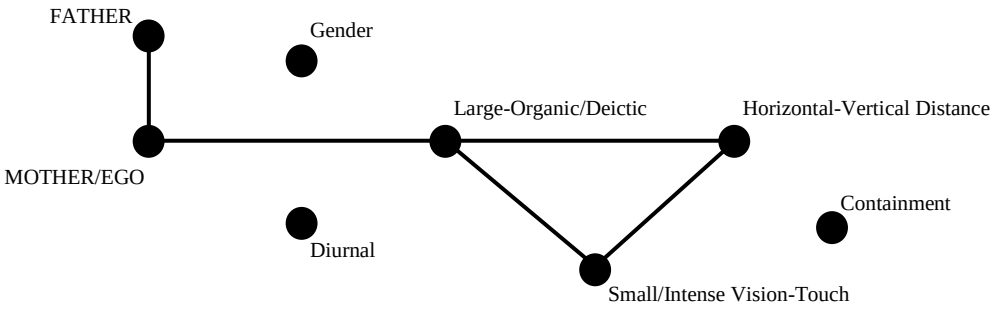

Figure 16: Relatively liberal suggestions for potential linguistic primitives and relations.

6. Conclusion

This thesis has investigated how cross-linguistic phoneme distributions of 56 fundamental oppositional concepts can reveal semantic relationships by looking into the actual linguistic forms of 75 genetically and areally spread out languages.

The questions asked were: Will oppositional word pairs of basic vocabulary show correlations between semantically related meanings solely based on their phonetic makeup? If such correlations are found, what is the phonetic makeup of each concept? If such correlations are found, which semantic domains and which relations between the semantic domains are found? Five different types of sound groupings were used dividing phonemes according to; the F2-frequency of vowels and the frequency of energy accumulation for consonants (Frequency); the sonority hierarchy (Sonority); a combination of the two aforementioned groupings (Combination); very general phonetic traits i.e. voiced, voiceless, vowels and consonants (General), and lastly, incorporating all traits of the four presented groupings (All). Through these sound groupings biplots were created, measuring how similar the phoneme distribution was among the investigated concepts. Also, phoneme distributions' over- and underrepresentation from the normal was calculated illustrating which sounds represented and lacked in each concept.

Based on the five biplots, it became evident that the relative positions of the concepts clusters were more or less the same, hence not too much can be said about which sound grouping revealed what semantic relationship when measured this way. In other words, the sound grouping based on Ohala's (1994) Frequency Code or the other four groupings did not differ

in what semantic clusters or relations were revealed based on phonetic makeup. What can be said however, is that the All-grouping yielded the clearest result which was attributed to the larger amount of data included than for the other four groupings.

Regarding actual results per concept, it was found that 41 out of the 56 concepts were judged significant according to at least one of the five sound groupings, and 13 of the 41 occurred in all of the sound groupings; MOTHER, I, FATHER, WIDE/BROAD, ABOVE, MAN, THIS, NARROW, SMALL, ROUGH, SHORT, BLACK, FLAT. The phoneme distribution-based clusters formed by the significant concepts displayed 10 semantic domains; *Small*, *Intense Vision-Touch*, *Large*, *Organic*, *Horizontal-Vertical Distance*, *Deictic*, *Containment*, *Gender*, *Parent*, *Diurnal*, and one lone concept; OLD. Further, several oppositional relations between concepts of different clusters, between concepts within the same cluster, as well as between whole clusters were found. Concerning the deviation from average phoneme distribution of the concept, many sound groups deviated over 50 %, and in some cases more than 100 % in both over- and underrepresentation. Still, some sound groups, including all the sound groups of the Combination-grouping, had to be excluded from the analysis due to their unreliability.

The concepts of embodiment together with phonosemantics were discussed as primary explanations for the results through bodily mimesis and neurological evidence for long-range connections between different areas of the brain, yielding not only connections between different meanings, but also between meanings and sounds. The oppositional relationships displayed between the concepts and clusters could be credited to the benefits of oppositional thinking in learning and language acquisition. The actual clustering of semantically related words could be attributed to possible common semantic origins; for example, that more general meanings preceding complex meanings could be based on adjective typology, fixed orders of lexicalization of perception verbs, ideophones and dimensional expressions, as well as that some dimensional adjectives are more easily learnt than others and that children start off by using words with fewer features.

Among the most important relationships found, the Large (e.g. LONG, DEEP) and Organic (e.g. ROUND, SOFT) concepts, and the Small (e.g. FEW, NARROW) and Intense Vision-Touch (e.g. DARK, HARD) concepts were found to occur entwined together, which suggests close relationships, possibly based on co-occurring characteristics, i.e. BIG and SMOOTH, SMALL and ROUGH etc. MOTHER and I sharing many phonological traits (especially

nasals) could be explained by the sound produced whilst breastfeeding and reinforced by the fact that the relaxed state of velum when producing such sounds could correspond to feeling secure, protected by the mother. FATHER having about as differing phoneme distribution as MOTHER and I, but differing in regard to actual phonemes could be explained by the infant wanting or needing to make a distinction between the two parental figures. The three-way distinction between the featured deictic concepts, I vs. THIS, THAT, YOU, HERE vs. THERE, corresponds well to the fact that over 90 % of the world's languages use two-way or three-way deictic systems, and correlates with studies concerning self-awareness. The semantic clusters containing the domains Small (possibly together with Intense Vision-Touch), Large-Organic and Horizontal-Vertical Distance form a tripartite opposition in the spatial dimension, which might correlate with very fundamental dimensions used for describing and understanding the world.

Several binary oppositional relations were found between the Small and Intense Vision-Touch, and Large-Organic clusters (e.g. SOFT-HARD. SMOOTH-ROUGH), probably because of their close connection to concrete sense perception, and perhaps also due to the fact that some of them have a common fundamental relationship based on the more general categories BIG and SMALL. However, a satisfactory explanation for LIGHT (not DARK) not occurring together with DARK and the other Intense Vision-Touch concepts was not found. Additionally, three pairs of oppositional concepts occurring closely together in the biplots based on phoneme distribution but also carrying similar semantic information (THIS-THAT, WOMAN-MAN, FULL-EMPTY) were suggested to be explained through a shared lexical core which was "phonosemantically inflected".

Finally, if all these clusters and connections are considered together, then eight possibly linguistic primitives materialize; FATHER, MOTHER/EGO, Large-Organic/Deictic, Small/Intense Vision-Touch, Horizontal-Vertical Distance, Gender, Diurnal and Containment.

What can be said with certainty is that solely based on the phonemes of the linguistic forms of 75 well-sampled languages, very evident semantic connections and relevant, often binary relationships between different semantic domains crystallized. Hence, this study illustrates the importance of meaning-meaning and sound-meaning connections, as well as of oppositional thinking constructing larger concept-networks, for human language.

Continuing on the present study could obviously include more languages in order to affirm the results further, including more concepts could yield more insight in what semantic clusters are affected by phonosemantics, but also which semantic domains and notions are connected within our minds. The results showing that oppositional relationships seem to be very fundamental lead to the conclusion that including more complex relations such as tastes and gradient scales of colors would probably be the next step as far as the chosen concepts are concerned. It could be beneficial to look into creating a more adequate sound classification system as well, which could incorporate the strengths of both the ASJP system of Wichmann et al. (2010) and the one featured in this study. Also, including consonant clusters, not separating affricates into plosives and fricatives and diphthongs into single vowels in a similar study as well as other ways of grouping sounds, except for Frequency, Sonority etc., could very well unlock other interesting connections between sound and meaning and also between meanings.

References

Abelin, Å. 1999. *Studies in Sound Symbolism.* Doctoral Dissertation: Gothenburg University.

Abelin, Å. & Allwood, J. 1984. *Tolkning av känsloprosodi - en kontrastiv studie*, unpubl. paper, Göteborg University, Department of Linguistics.

Abrahamsson, N. 2001. *Acquiring L2 Syllable Margins. Studies on the simplification of onsets and codas in interlanguage phonology*. Doktorsavhandling, Centrum för tvåspråkighetsforskning, Stockholms universitet.

Abrahamsson, N. 2003a. Universal constraints in L2 coda production: the case of Chinese/Swedish interphonology. In: Costamagna, L. & Giannini, S. (eds.) *La fonologia dell'interlingua. Principi e metodi di analisi*. Milano: Franco Angeli.

Abrahamsson, N. 2004. Fonologiska aspekter på andraspråksinlärning och svenska som andraspråk. In: Hyltenstam Kenneth & Lindberg Inger (eds.) *Svenska som andraspråk: i forskning, undervisning och samhälle*. Lund: Studentlitteratur.

Ahlner, F & Zlatev, J. 2010 Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign System Studies*, 38(1/4): 298-348.

Barlett, E. J. 1976. Sizing things up: the acquisition of the meaning of dimensional adjectives. *Journal of Child Language 3*, 205-219.

Robert Beekes (with the assistance of Lucien van Beek) 2013. "φλέγω" in: *Etymological Dictionary of Greek*. Indo-European Etymological Dictionaries Online. Edited by Alexander Lubotsky. Brill. Brill Online. November 16, 2013.

Bentley, M. & Varon, E. 1933. An accessory study of phonetic symbolism *Amer. J. Psychol.* 45, 76-86.

Berlin, B. & Kay, P. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkley, CA: University of California Press.

Bickerton, D., 1990. *Language and species*. Chicago: University of Chicago Press.

Bloomfield, M. 1895. On assimilation and adaptation in congeneric classes of words. *Am. J. Phil.16*, 409–434.

Bohnemeyer, J. 2003. NSM withough th Strong Lexicalization Hypothesis. *The Theoretical Linguistics 29*.

Bowerman, M. (1996). The origins of children's spatial semantic categories: Cognitive vs. linguistic determinants. In J. J. Gumperz, & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 145-176). Cambridge University Press.
Brown, C. H., Holman, E. W., Wichmann, S., Velupillai, V. 2008. Automated classification of the world's languages: a description of the method and preliminary results. *STUF – Language Typology and Universals* 61.4: 285–308.
Brown, P. & Levinson, S. C. 1993. 'Recentering' in Mayan spatial description, with special reference to Tzeltal. *Journal of Linguistic Anthropology, 3*(1), 46-74.
Brownell C. A, Zerwas S., & Ramani G. B. 2007. "So big": The development of body selfawareness in toddlers. *Child Development* 78.
Bybee, J., Perkins, R. & Pagiuca, W. 1994. *The Evolution of Grammar*. Chicago IL: The University of Chicago Press.
Chastaing, M. 1958. "Le symbolisme de voyelles: significations des "i" I& II", *Journal de Psychologie* 55, 403-423 & 461-481.
Chastaing, M. 1965. "Pop - fop - pof - fof", *Vie et Langage* 159, 311-317.
Chastaing, M. 1966. "Si les *r* étaient des *l*", (Part 1) *Vie et langage* 173, 468-472; (Part 2) 174, 502-507.
Cinque, G. 2013. Cognition, typological generalizations, and Universal Grammar. In *SI: Syntax and cognition: core ideas and results in syntax, Lingua 130* ed. Luigi Rizzi, 50-65.
Clark, E. V. 1973. What's in a word?: On the child's acquisition of semantics in his first language. In *Cognitive development and acquisition of language*. Moore, Timothy E (ed.). New York: Academic Press.
Clark, H. H. 1973. Space, time, semantics, and the child. In: Moore, Timothy E. *Cognitive development and the acquisition of language*. Oxford: Academic Press.
Collins, A. M. & Quillian M. R. 1969. Retrieval time from semantic memory: *Journal of Verbal Learning and Verbal Behaviour* 8: 240-8.
Cytowic, R. 1989. *Synaesthesia - a union of the senses*. New York: Springer Verlag.
Deese, J. 1965. *The structure of associations in language and thought*. Baltimore: The John Hopkins Press.
de Villiers, J. G., & de Villiers, P. A. 1978. *Language acquisition*. Cambridge, MA: Harvard University Press.
Diessel, H. 2005. Distance contrasts in demonstratives. In *World atlas of language structures*, eds. M. Haspelmath, M. Dryer, D. Gil & B. Comrie, 170-173. Oxford: Oxford University Press.
Diffloth, G. 1994. i: big, a: small. In: L. Hinton, J. Nichols, J.J. Ohala (eds). *Sound symbolism*. Cambridge: Cambridge University Press, 107-114.
Dingemanse, M. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass, 6*(10), 654–672.
Dixon, Robert M. W. 1982. *Where have all the adjectives gone?* In: Dixon, Robert M.W. 1982 (ed.) 1982. Where have all the adjectives gone? and other essays in Semantics and Syntax. Amsterdam: Mouton. 1-62.
Dixon, R. M. V. 2010. Basic Linguistic Theory. Volume 1 Methodology. Oxford University Press.
Donald, M. 1991. *Origins of the Modern Mind. Three Stages in the Evolution of Culture and Cognition*. Cambridge, Mass.: Harvard University Press.
Everett, D. L. 2005a. Periphrastic pronouns in Wari'. *International Journal of American Linguistics, 71:3,* pp 303-326.
Fónagy, I. 1963. *Die metaphern in der phonetik*. The Hague
Geeraerts, D. 2010. *Theories of Lexical Semantics*. Oxford: Oxford University Press.

Gibson, J. J. 1977. The Theory of Affordances. In *Perceiving, Acting, and Knowing*. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing*. Hillsdale, New Jersey: Erlbaum.
Goddard, C. & Wierzbicka, A. (eds.). 2002. *Meaning and Universal Grammar: Theory and Empirical Findings* (2 volumes). Amsterdam/Philadelphia: John Benjamins.
Gussenhoven, C. 2001. Intonation and biology. In Jacobs, Haike & Wetzels, W.Leo (Eds.), *Liber Amicorum Bernard Bichakjian*. Maastricht: Shaker. 59-82.
Hale, K. 1971. A note on a Walbiri tradition of antonymy. D. Steinberg & L. Jacobovits (eds.) *Semantics. An interdiciplinary reader in philosophy, linguistics and psychology*. Cambridge: Cambridge University Press.
Hamlin, K., Wynn, K. & Bloom, P. 2007. Social evaluation by preverbal infants. *Nature, 450,* 557–559.
Haspelmath, M. & Tadmor, U. 2009. The Loanword Typology project and the World Loanword Database. In: Haspelmath, Martin & Tadmor, Uri (eds.) *Loanwords in the World's Languages: A Comparative Handbook*. Berlin: Mouton de Gruyter, 1-34.
Heine, B. & Kuteva, T. 2007. *The Genesis of Grammar: A Reconstruction (Studies in the Evolution of Language)*. Oxford University Press, Oxford.
Herder, J. G. 1772/1969. Essay on the origin of speech in P.H. Salus (ed.) *On language: Plato to von Humboldt, Holt, Rinehart and Winston*. New York, pp.147-166.
Hinton, L., Nichols, J., & Ohala, J. 1994. *Sound Symbolism*. Cambridge: Cambridge University Press.
Hrdy, S. 2011. *Mothers and others: the evolutionary origins of mutual understanding*. 1st Harvard University Press paperback ed. Cambridge, Mass.: Belknap Press of Harvard University Press.
Hogg, R. M. & McCully C. B. 1987. *Metrical Phonology: A Course Book*. Cambridge: Cambridge University Press.
Holman, E., Wichmann, S., Brown, C. H., Velupillai, H., Müller, A. & Bakker, D. 2008. Explorations in automated language classification. *Folia Linguistica* 42(2): 331-354.
Jakobson, R., Waugh, L. 1979. *The sound shape of language*. Sussex: Harvester Press.
Jespersen, O. 1922. *Language - its nature, development and origin*. London: Allen and Unwin.
Johansson, N. & Carling, G. (submitted). *The De-Iconization and Rebuilding of Non-Arbitrariness in Spatial Deixis: A Indo-European Case Study*.
Johansson, N. & Zlatev, J. (2013). Motivations for Sound Symbolism in Spatial Deixis: A Typological Study of 101 languages. *The Public Journal of Semiotics*.
Johnson, K. 2008. *Quantitative methods in linguistics*. Malden, Mass.: Blackwell
Justeson, J. & Katz, S. 1991. Co-occurrence of antonymous adjectives and their contexts. *Computational Linguistics* 17:1, pp 1-19.
Kemmerer, D. 1999. "Near" and "far" in language and perception. *Cognition*, *73*, 65–73.
Kita, S., Kantartzis, K., and Imai, M. 2010. Children learn sound symbolic words better: Evolutionary vestige of sound symbolic protolanguage. In: Andrew D. M. Smith, Bart De Boer, Marieke Schouwstra, *Evolution Of Language, The - Proceedings Of The 8Th International Conference (Evolang8).* Singapore: World Scientific Publishing Company Co. Pte. Ltd., 206-213.
Kramer, M. S. & Kakuma, R. 2009. Optimal duration of exclusive breastfeeding. *The Cochrane Library*, Issue 1.
Köhler, W. 1929 *Gestalt Psychology*. New York: Liveright.
Ladefoged, P. 2001. *Vowels and consonants: An introduction to the sounds of language*. Oxford: Blackwell Publishers.
Lakoff, G. & Johnson, M. 1980. *Metaphors we live by*. Chicago: Chicago University Press.

Lakoff, G. 1987. *Women, Fire and Dangerous Things*: *What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
Landau, B. & Jackendoff, R. 1993, '"What" and "Where" in Spatial Language and Spatial Cognition'. *Behavioral and Brain Sciences 16*, 217-238.
LaPolla, R. J. 1995. An Experimental Investigation into Sound Symbolism as it Relates to Mandarin Chinese. *Sound Symbolism*, ed. by Leanne Hinton, Johanna Nichols, and John Ohala, 130-147. Cambridge: Cambridge University Press.
Lévi-Strauss, C. 1955. The Structural Study of Myth. In T. A. Sebeok (ed.*) Myth: A Symposium*. Bloomington, IN, pp. 81-106.
Lévi-Strauss, C. 1984. The Story of Asdiwal. In Alan Dundes (ed.) *Sacred Narratives: Readings in the Theory of Myth*. Berkeley, CA, pp. 295-314.
Li, C. N. & Thomson S. A. 1977. A Handbook in Comparative Tai. Honolulu: University Press of Hawaii.
Li, C. N. & Thomson S. A. 1981. *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.
Lieberman P., Blumstein S. E. 1988. *Speech physiology, speech perception, and acoustic phonetics*. Cambridge: Cambridge University Press.
Linell, P. 1982. Speech errors and grammatical planning of utterances. Evidence from Swedish. W. Koch, C. Platzack & G. Tottie (eds.) *Textstrategier i tal och skrift*. Stockholm: Almqvist & Wiksell.
Lundbladh, C.-E. 1988. Adjektivets komparation i svenskan. En semantisk beskrivning. *Lundastudier i nordisk språkvetenskap* A 40. Lund: Lund University Press.
Mace, W. M. 1977. James J. Gibson's strategy for perceiving: Ask not what's inside your head, but what your head's inside of. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing*. Hillsdale, New Jersey: Erlbaum.
Malle, B. F. 2002. The relation between language and theory of mind in development and evolution. In T. Givon, B. F. Malle (Eds.), *The evolution of language out of pre-language* (pp. 265–284). Amsterdam: Benjamins.
McTurk, R. (ed.) 2005. *A Companion to Old Norse-Icelandic Literature and Culture*. Oxford: Blackwell.
Medin, D. L. & Brian H. R. 1992. *Cognitive Psychology*. San Diego, CA: Harcourt Brace Jovanovich.
Moore, C. 2006. *The development of commonsense psychology*. Mahwah, NJ: Lawrence Erlbaum Associates.
Müller, M. 1861. *Lectures on the Science of Language*. London: Longmans, Greens and Co.
Müller, A. L. 1960. *Experimentelle Untersuchungen zur stimmlichen Darstellung von Gefühlen*, Dissertation. Göttingen.
Muysken, P. 2009. Functional categories: an evolutionary perspective. In Botha, Rudi & Henriëtte de Swart (eds.) *Language evolution: the view from restricted linguistic systems*. LOT Occasional series 10.
Nakagawa, H. 2012. The importance of TASTE verbs in some Khoe languages. *Linguistics* 50–3, 395–420.
Needham, A., & Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition, 47*, 121-148.
Newman, S. 1933. Further experiments in phonetic symbolism, *American Journal of Psychology* 45, 53-75.
Nichols, J. 1992. *Linguistic Diversity in Space and Time.* Chicago: University of Chicago Press.

Oberman, L. A., & Ramachandran, V. S. (2008). Preliminary evidence for deficits in multisensory integration in autism spectrum disorders: The mirror neuron hypothesis. *Social Neuroscience, 3*(3–4), 348–355.
Ohala, J. J. 1994. The frequency code underlies the sound-symbolic use of voice pitch. In: L. Hinton, J. Nichols, J.J. Ohala (eds.) *Sound symbolism*. Cambridge: Cambridge University Press, 325-347.
Osgood, C. E., Suci, G. J. & Tannenbaum P. H. 1957. *The measurment of meaning,*
University of Illinois Press. Urbana, Illinois
Pagel M., Atkinson Q. D. & Meade A. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449:717–720.
Pagel, M., Atkinson, Q. D., Calude, A. S. & Meade, A. 2013. Ultraconserved words point to deep language ancestry across Eurasia. *Proceedings of the National Academy of Sciences*. Published online before print May 6 2013.
Paradis, C., Willners, C. & Jones, S. 2009. Good and bad opposites. Using textual and experimental techniques to measure antonym canonicity. *The Mental Lexicon* 4:3, 380–429. Amsterdam: John Benjamins Publishing Company.
Plato (c. 428-348 B.C.) *Cratylos dialogue* 1962. Loeb Classical Library. Cambridge: Harvard University Press.
Premack, A., & Premack, D. 1983. *The mind of an ape*. New York: Norton.
Pulvermüller, F. 1999. *Words in the brain's language*. Behavioral and Brain Sciences 22(2)253–336.
Ramachandran, V. S. & Hubbard, E. M. 2001. Synaesthesia – a window into perception, thought and language. *Journal of Consciousness Studies* 8: 3–34.
Reay, I. E. 1994. Sound Symbolism. In R.E. Asher (ed.), The *Encyclopedia of Language and Linguistics, Vol. 8*, 4064-70. Oxford: Pergamon Press.
Saeed, J. 2003. *Semantics, Second Edition*. Malaysia: KHL Printing Co Sdn Bhd.
Saussure, F. 1916/1983. *Course in general linguistics*, Duckworth. London.
Sapir, E. 1929. A study in phonetic symbolism. In D. Mandelbaum (ed.) *Selected writings*. Berkeley. California.
Sereno, J. A. 1994. Phonosyntactics. In: L. Hinton, J. Nichols, J.J. Ohala (eds). *Sound symbolism*. Cambridge: Cambridge University Press, 263-275.
Shimotori, M. 2013. Conceptual Contrasts: *A Comparative Semantic Study of Dimensional Adjectives in Japanese and Swedish*. Doctoral Dissertation: Umeå Studies in Language and Literature 17.
Shinohara, K. & Kawahara, S. (2012). A cross-linguistic study of sound symbolism: The images of size. In *Proceedings of BLS 36*. Berkeley: BLS
Short, T. L. 2007. *Peirce's theory of signs*. New York: Cambridge University Press.
Slobin, D. I. 1997. Mind, code and text. In *Essays on language function and language type*. Bybee, J. Haiman, J, & Thompson, S., (eds.), 437-467. Philadelphia: John Benjamins.
Slobin, D. I. 2000. Verbalized Events, in *Evidence for linguistic relativity*. Niemeier, S, & Dirven, R (eds), 107-137. Amsterdam: Benjamins.
Swadesh, M. 1971. *The origin and diversification of language*. Edited post mortem by Joel Sherzer. London: Transaction Publishers.
Söderpalm, E. 1979. Speech errors in normal and pathological speech. *Travaux de l'institut de linguistique de Lund* 14. Lund: CWK Gleerup.
Talmy, L. 1983. How language structures space. In: *Spatial orientation: Theory, research, and application,* ed. H. Pick & L. Acredelo. Plenum Press.
Tomasello, M. 1992 *First Verbs: A Case Study of Early Grammatical Development.* Cambridge: Cambridge University Press.

Traunmüller, H. 1994. Sound symbolism in deictic words. In: Hans Auli and Peter af Trampe (eds.) *Tongues and Texts Unlimited. Studies in Honour of Tore Jansson on the Occasion of his Sixtieth Anniversary*. Dept. of Classical Languages, Stockholm University, 213-234.
Tsuru, S., Fries, H. S. 1933. A problem in meaning. *Journal of Gen*. Psychology 8, 281-284
Ultan, R. 1978. Size-sound symbolism. In: J. Greenberg (ed.), *Universals of human language 2*.
Usnadze, D. 1924. Ein experimenteller Beitrag zum problem der psychologischen Grundlagen der Namengebung, *Psychologische Forschung* 5, 24-43.
Veselinova, L. N. 2005. *Suppletion in Verb Paradigms*. Amsterdam: John Benjamins Publishing Company.
Viberg, Å. 1984. The verbs of perception: a typological study. *Linguistics* 21(1). 123–162.
Viberg, Å. 2001. Verbs of perception. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, 1294 –1309. Berlin & New York: Walter de Gruyter.
Wescott, R. W. 1971. *Linguistic Iconism*. Language 47: 4 16-428.
Whorf, B. L. 1956/1967. *Language, thought and reality: selected writings,* ed J.B. Carroll (ed), The MIT paperback series 5. Cambridge: MIT Press.
Wichmann, S., Holman, E. W. & Brown, C. H. 2010. Sound Symbolism in Basic Vocabulary. *Entropy 12*, no. 4: 844-858.
Wiendold, G. & Rohmer, U. 1997. On Implications in Lexicalizations for Dimnesional Expressions. In: Yamanaka, Kei & Ohori, Toshio. (eds.) *The locus of meaning. Papers in honor of Yoshihiko Ikegami*. Tokyo: Kuroshio publishers.
Wierzbicka, A. 1972. *Semantic Primitives*. Frankfurt: Athenäum.
Wilde, K. 1958. 'Naive und kunstlerische Formen des graphischen Ausdrucks', *Ber. 21. Kongr. Dtsch. Ges. Psychol*., Göttingen: Verlag fur Psychologie, Bonn, pp. 157-159.
Willners, C. 2001. *Antonyms in Context: A Corpus-based Semantic Analysis of Swedish Descriptive Adjectives*. Lund University.
Wisseman, H. 1954. *Untersuchungen zur Onomatopoie*, Part I: Die sprachpsychologischen Versuche, Winter. Heidelberg.
Woodworth, N. L. 1991. Sound symbolism in proximal and distal forms. *Linguistics 29*: 273-299.
Yates, F. A. 1966. *The Art of Memory*. Chicago: University of Chicago Press.
Yue-Hashimoto, A. O. 1980. Word play in language acquisition: a Mandarin case. *Journal of Chinese Linguistics* 8(2): 181-204.
Zlatev, J. 2005. What's in a schema? Bodily mimesis and the grounding of language. In *From Perception to Meaning: Image Schemasin Cognitive Linguistics*, B. Hampe (ed.), 323-342. Berlin: Mouton.
Zlatev, J., Persson, T. & Gärdenfors, P. 2005. *Bodily Mimesis as "the Missing Link" in Human Cognitive Evolution*. LUCS 121. Lund: Lund University Cognitive Studies.

Language Sources

Australian Society for Indigenous Languages: Warlpiri. (http://203.122.249.186/Lexicons/Warlpiri/index-english/index.htm) data collected 2013-02-20.
Bambara On-line Lexicon(http://www.bambara.org/lexique/index-english/main.htm) data collected 2013-02-14.
Blust, Robert 2003. *Thao Dictionary*. Language and Linguistics Monograph series number A5. Institute of Linguistics, Academica Sinica. Taipei, Taiwan.

Bugaeva, Anna & Endo, Shiho (eds.). A talking dictionary of Ainu: a new version of Kanazawa's Ainu conversational dictionary with recordings of Mrs Setsu Kurokawa. Funded by ELDP, Hans Rausing Endangered Languages Project. (http://lah.soas.ac.uk/projects/ainu/dict.html) data collected 2013-02-20.
Cambridge Dictionaries Online: Turkish.(http://dictionary.cambridge.org/dictionary/turkish/) data collected 2013-03-6.
C'Lela—English—Hausa Dictionary. (http://homepage.ntlworld.com/david.rowbory/clela/) data collected 2013-02-14
Ethnologue Online. (http://www.ethnologue.com/) 2013-12-11
Fataluku language Project (http://www.fataluku.com/dictionary/search/) data collected 2013-02-21.
fie.nipa Dictionaries: Kinyarwanda. (http://words.fienipa.com./all/ee?t=smooth&ff=en&ft=rw) data collected 2013-03-7.
Free Personal Pronoun System database. (http://languagelink.let.uu.nl/fpps/) data collected 2013-03-8.
Free English to Korean Dictionary. (http://www.koreandictionary.net/search?search=narrow&searchtype=english) data collected 2013-03-8.
Gambill, G. T. Cheyenne-English and English-Cheyenne online dictionary. (http://www.freelang.net/online/cheyenne.php?lg=gb) data collected 2013-02-21.
Inkuktitut Living Dictionary. (http://www.livingdictionary.com/term/viewTerm.jsp?term=49170249693) data collected 2013-03-8.
isiZulu.net. (http://isizulu.net/) data collected 2013-02-14.
Kahn, L. & Levi, Y. (eds) 1995. *Oxford English - Hebrew/Hebrew - English Dictionary*. Jerusalem: Kernerman.
Khoekhoe etymology. Compiled by George Starostin. (http://starling.rinet.ru/kkhet.pdf) data collected 2013-02-14.
Li, Paul Jen-kuei & Tsuchida, Shigeru 2006. *Kavalan Dictionary*. Language and Linguistics Monograph series number A-19. Institute of Linguistics, Academica Sinica. Taipei, Taiwan.
Paas, Steven. Chichewa/Chinyanja Dictionary Project. The Fourth Edition. (http://translate.chichewadictionary.org/) data collected 2013-03-7.
Peace Corps: Fula Language courses. (http://resourcepage.gambia.dk/langabot.htm) data collected 2013-03-6.
Rosetta Project. (http://archive.org/search.php?query=swadesh%20collection:rosettaproject) data collected 2013-03-7.
Sesotho Online. (http://bukantswe.sesotho.org/display.php?action=search&word=thick&type=full) data collected 2013-02-14.
The Intercontinental Dictionary Series. (http://lingweb.eva.mpg.de/ids/) data collected 2013-03-7.
The Online English-Tibetan Dictionary (http://www.eng-tib.com/) data collected 2013-02-21.
Thieberger, Nicholas (2011-05-05) Dictionary of South Efate. (http://paradisec.org.au/SELexicon/index-english/main.htm) data collected 2013-02-20.
Tok Pisin/English Dictionary compiled by Barhorst, T. D. & O'Dell-Barhorst, S. (http://www.june29.com/HLP/lang/pidgin.html) data collected 2013-02-21.
Wolff, J. U. 1972. Dictionary of Cebuano Visayan. (http://www.bohol.ph/wced.php?sw=thick&lang=All&hw=1) data collected 2013-02-21.

World Loanword Database. (http://wold.livingsources.org/) data collected 2013-03-8.

Appendix A: Linguistic Forms.

The featured 75 languages and 56 concepts of the study. Lingusitic forms are ordered according to language family or language group and converted into IPA (somehwhat simplified).

	Afro-Asiatic				Austronesian											
	Hausa	**Hebrew**	**Iraqw**	**Tarifiyt Berber**	**Hawaiian**	**Tongan**	**Rotuman**	**Malagasy**	**Takia**	**Cebuano**	**South Efate**	**Tetum**	**Malay**	**Thao**	**Kavalan**	**Seediq**
I	ni	ani	aniŋ	nəʃ	au	ou	ŋou	aho	ŋai	ikaw	kineu	haʔu	aku	jakuʔ	iku	jaku
YOU	kai	ata	kuŋ	ʃək	ʔoe	koe	ʔæe	ianao	oŋ	aku		o	kamu	ihu	ʔaisuʔ	isu
BIG	baba	gadol	ur	amɔqran	nui	lahi	tiʔu	be	tbu	daku	top	bot	besar	maraʔin	raja	paru
SMALL	karami	ktan	nina	amɔʕnan	iki	siʔisiʔi	meaʔmeʔa	keli	tsa	dijut	rik	kik	ketʃil	laqqikuʔaj	kitut	bitsiq
GOOD	mai kjau	tov	ħoʔ	ʃən	maikaʔi	lelei	lelei	tsara	uja	maj	wi	diʔak	baiʔ	du	nŋi	malu
BAD	maras kjau	ʁa	tlakw	aʕəfan	maikai ʔole	kovi	raksaʔa	ratsi	sae	amat	mkal			maqarman	sukaw	naqah
THIS		ze	ti	a	keia			iti	jen	kini	ne	ida	ini	inaj	zau	ni
THAT		ze	siŋ	in	kela			iri	an	kana	tego	ida neʔe	itu	haja	naj	gaga
MUCH/MANY	jawa	ʁabim	ur	atas	nui	tokolahi	tiʔu	betsaka	wei	daghan	lap	barak	banjaʔ	manaʃa	waza	egu
FEW	kadan	meʔat	peraj	ðrus	kakaʔikahi	tokosiʔi	meaʔmeʔa	vitsi	tata	minus	nrfal			ladadu	kia	tikuh
BEFORE	kaɸin	lifnej	gera	qβər i	mua	ʔi muʔa	mumua	aloha		adisir	tetwei	uluk	sebelum	kahiwan	masaŋ	berah
AFTER	bajan	axarej	alu	awan i	hope	hili	fakmuri	aoriana	mar na		ntakun	hafoin	kemudian		tiziw	bobo
ABOVE	a sama	lmala	gawa	s nɔz i	maluna	ʔoluŋa	ʔe rere	amboni	fo na		elag		di atas	ifafaw	babaw	bobo
BELOW	a kasa	mitaxat	gamu	adu	lalo	ʔi lalo	sio	ambani	paen na	ubus			di bawah	iproq	ʁbeŋ	truma
FAR	nesa	ʁaxok	saw	gwɔz	mamao	mamaʔo	sousou	laviŋa	asaw	atua	emae	dok	jauh	ihaziʃ	laul	thejaq
NEAR	kusa da	kaʁov	tsew	aðɔs	kokoke	ofi	ʔelʔele	akaiki	smeik	pidpid		besik	dekat	eqoał	zaki	daliŋ
WOMAN	mace	iʃa	ʕameni	tamətut	wahine	fefine	hɔni	vehivavi	pein	babae	nmatu	feto	perempuan	binanauʃað	tazuŋan	mqedin
MAN	namid͡ʒi	iʃ	ħawata	ataras	kanaka	taŋata lahi	fa	lehilahi	tamol	lalake	kano			ajuði	ʁunanaj	rseno
WHITE	fari	lavan	ʕawak	aʃɔmrar	keʔokeʔo	hinehina	fisi	fotsi	malkouk	puti	tar	mutin	putih	mapuði	busaʁ	bhege
BLACK	baki	ʃaxoʁ	boʕ	aβarʃan	ʔeleʔele	ʔuliʔuli	kele	mainti	tdomtbun	itum	got	metan	hitam	maqosum	tŋen	qalux
HOT	zaɸi	xam	daʕaʕam	aɣ	wela	vela	sunu	mahamai	wanana	init	ftin	manas	panas	mahnar	skwaru	mtilux
COLD	sanji	kaʁ	tsaʔ	asəmað	anu	momoko	matiti	maŋgatsiaka	barum	bugnaw	mlanr	malirin	dingin	mahaðiwhiw	sen	msekuj
HERE		po	diri	ða	ʔaneʔi			eti	je	adia	sa	iha neʔe	di sini	inaj	zi	hini
THERE		ʃam	didaʔ	ðin	laila			eri	a	atua	san	iha neʔeba	di sana	isahaj	tajan	hija
LONG	t͡saɸo	aʁok	tɬer	azijra	loa	loloa	roa	lava	mlae	tas	naframwen	naruk	panjaŋ	maqolijus	ʁruŋ	qnedis
SHORT	gad͡ʒere	katsar	ququmar	aquðað	pokole	nounou	luka	fohi	katka	mubo	mit	badak	pendek	luiʃ	kzu	blebu

	Afro-Asiatic				Austronesian											
	Hausa	**Hebrew**	**Iraqw**	**Tarifiyt Berber**	**Hawaiian**	**Tongan**	**Rotuman**	**Malagasy**	**Takia**	**Cebuano**	**South Efate**	**Tetum**	**Malay**	**Thao**	**Kavalan**	**Seediq**
NIGHT	dare	lajla	xwera	dʒirət	po	poʔuli	poŋi	alina	tdom	gabil	pog	kalan	malam	mahumhum	ʁabi	keman
DAY	rana	jom	bal	nha	la	ʔaho	terɔni	aŋdo	adjan	adlaw	mau naliati	loron	hari	qali	lan	dijan
FULL	cika	male	hats	ʃua	piha	fonu	hoi	feno	awan isai	punu	tisoksok			mabaʃbaʃ	bisuʁ	steŋe
EMPTY	bakome a ciki	ʁejk	kahar	xwa	haka	maha	tafa	foana	kao	hawaŋ	pal	kaʔare	kosoŋ	aðað		
NEW	sabo	xadaʃ	ʕaben	zðið	hou	foʔou	foʔou	vaovao	fou	bagu	foum		baru	faqhu	tasu	bgurah
OLD	tsoho	jaʃan	qaren	amɔqran	ʔaoʔo	motuʔa	mafua	antiŋa	wagama	dan	motu	katuas	tua	sasað	ʔzan	tsmutsats
ROUND	kewajaje	agol	gumbalalaħ	aquraj	poepoe	fuopotopoto	kɔlkɔlu	boribori	tlanti	bukud	manopnop			mabuðoq	turun	mtumun
FLAT	bai daja		peħ	wata	palahalaha	lafalafa	teaptepa	lemaka	labaka	lapad	matit		litʃin	mabareð	tpajas	qepi
DRY	busaʃe	javeʃ	kahar	azəɣ	maloʔo	momoa	mamasa	maina	gos	laja	gar	maran	keriŋ	makutbað	iseŋ	mdeŋu
WET	ʤikake	ʁatov	naʕ	uf	maʔu	hauhau	matmata	lena	gan	basa	sumsmo		basah	matuðu	zizi	mhuriq
WIDE/BROAD	faɗi	ʁaxav	intlaxw	miriw	laula	laulahi	tafa	malalaka	baŋan tbun	lawaŋ	polplo		lebar	mabaram	tabaj	glahaŋ
NARROW	maras faɗi	tsar	iiraakw	ħsa	laʔiki	fasiʔi	ʔele	eti	pŋpoŋa	higpit		klot	sempit	maqtit	niku	dgehin
THICK	kauri	ave	diqi ur	gða	manoanoa	matolu	mafolu	matevina	mtnol	baga	matol	mahar	tebal	makuʃtor	ʁutuz	qtehul
THIN	siriri	ʁaze	diqi nina	azðað	wiwi	pahauhau	mahini	manifi	mrjasa	nipis	mrara	krekas	kurus	lagqesusaj	impis	rqeliŋ
SMOOTH	sumul	xalak	motsotsoq	arəqaɣ	malino	hamolemole	marmarɔri	malama	smut	hamis	mal	kaber	halus	madeqro	laziu	mtbale
ROUGH	kauʃi	mxuspis	xaslasla?	ahaʃaw	hoʔolua	petepete	vɔnu	marao	raf	gansal	liuliu				nisan	qtehul
HEAVY	nauji	kaved	iloʔ	ðqər	kaumaha	mamafa	maha	mavesaŋa	mrwa	bugat	mten	todan	berat	mabrek	zineq	tshedin
LIGHT (not HEAVY)	maras nauji	kal	inslaħ	fsəs	mama	maʔamaʔa	tʃeamtʃema	maivana	pasama	gan	mram	kman	riŋan		ineŋat	tslokah
DARK	duhu	xaʃuk		ħðəq	maku?e	fakapoʔupoʔuli	maksulu	antiŋa		sulup	nmalko		malam	mahumhum	kumit	kuŋ
LIGHT (not DARK)	haske	vahiʁ		aʃəmrar	halakea	mama	maʔa	tanora		hajag	sor				tum	ledax
FAST	sauri	mahiʁ	ganslaj	s tazra	ʔawiwi	vave	væve	haiŋgana	wlwalemi	kusaj		lalais	tʃepat	mabiskaw	kramkamut	knuwa
SLOW	maras sauri	iti	tsegis	ʃwaj ʃwaj	lohi	tuai	fepi	miadana	pasak nami	hinaj	mailum	neineik	lamban	miaqawan	luŋut	knhuwe
HARD	tauri	xazek	gawid	qsəħ	ʔoʔoleʔa	fefeka	moumou	mafi	sakar	tiga	kerkerai			makuʃrak	qatmu	sadux
SOFT	lauʃi	ʁak	wanana?	aðəβ	palupalu	molu	maru	malemi	matala	humuk	mailum		lembut	mimbulnu	puʁu	mhenuk
DEEP	zurfi	amuk	tsalʔi	adʒəɣ	hohonu	loloto	lala	lalina	kror	lalum		as	tiŋgi	marukruk		truma
SHALLOW	maras zurfi	ʁadud	parampeħ	aʕra	papaʔu	mamaha	ʔelʔele	marivo	ftfota	mabaw			daŋkal	makapa	itaza	
HIGH	bisa	gavua	gawa	aʕra	kiʔekiʔe	maʔoluŋa	lamlama	amboni	lak na	habug	elag	as	tiŋgi		ibabaw	baro
LOW	fasa	gamuk	baraj	aða	lalo	maʔulalo	ʔe	ambani	tan na	bahadu		badak	pendeʔ	luiʃ		blebu
MOTHER	uwa	ima	ajo	jəma	makuahine	faʔe	oʔhoni	reni	tna	nanaj	raiten	inan	ibu	ina	na	bubu
FATHER	uba	av	baba	βaβa	makua kane	tamai	øʔfa	rai	tama	itaj	tata	aman	bapaʔ	ama	tama	tama

	Afro-Asiatic				Austronesian											
	Hausa	**Hebrew**	**Iraqw**	**Tarifiyt Berber**	**Hawaiian**	**Tongan**	**Rotuman**	**Malagasy**	**Takia**	**Cebuano**	**South Efate**	**Tetum**	**Malay**	**Thao**	**Kavalan**	**Seediq**
I	ni	ani	aniŋ	nəʃ	au	ou	ŋou	aho	ŋai	ikaw	kineu	haʔu	aku	jakuʔ	iku	jaku
YOU	kai	ata	kuŋ	ʃək	ʔoe	koe	ʔæe	ianao	oŋ	aku		o	kamu	ihu	ʔaisuʔ	isu
BIG	baba	gadol	ur	amɔqran	nui	lahi	tiʔu	be	tbu	daku	top	bot	besar	maraʔin	raja	paru
SMALL	karami	ktan	nina	amɔʕnan	iki	siʔisiʔi	meaʔmeʔa	keli	tsa	dijut	rik	kik	ketʃil	laqqikuʔaj	kitut	bitsiq
GOOD	mai kjau	tov	hoʔ	ʃən	maikaʔi	lelei	lelei	tsara	uja	maj	wi	diʔak	baiʔ	du	nŋi	malu
BAD	maras kjau	ʁa	tɬakw	aʕəfan	maikai ʔole	kovi	raksaʔa	ratsi	sae	amat	mkal			maqarman	sukaw	naqah
THIS		ze	ti	a	keia			iti	jen	kini	ne	ida	ini	inaj	zau	ni
THAT		ze	siŋ	in	kela			iri	an	kana	tego	ida neʔe	itu	haja	naj	gaga
MUCH/MANY	jawa	ʁabim	ur	atas	nui	tokolahi	tiʔu	betsaka	wei	daghan	lap	barak	banjaʔ	manaʃa	waza	egu
FEW	kadàn	meʔat	peraj	ðrus	kakaʔikahi	tokosiʔi	meaʔmeʔa	vitsi	tata	minus	nrfal			ladadu	kia	tikuh
BEFORE	kaɸin	lifnej	gera	qβɔr i	mua	ʔi muʔa	mumua	aloha		adisir	tetwei	uluk	sebelum	kahiwan	masaŋ	berah
AFTER	bajan	axarej	alu	awan i	hope	hili	fakmuri	aoriana	mar na		ntakun	hafoin	kemudian		tiziw	bobo
ABOVE	a sama	lmala	gawa	s nɔz i	maluna	ʔoluŋa	ʔe rere	amboni	fo na		elag		di atas	ifafaw	babaw	bobo
BELOW	a kasa	mitaxat	gamu	adu	lalo	ʔi lalo	sio	ambani	paen na	ubus			di bawah	iproq	ʁbeŋ	truma
FAR	nesa	ʁaxok	saw	ɡwəz	mamao	mamaʔo	sousou	laviŋa	asaw	atua	emae	dok	jauh	ihaziʃ	laul	thejaq
NEAR	kusa da	kaʁov	tsew	aðɔs	kokoke	ofi	ʔelʔele	akaiki	smeik	pidpid		besik	dekat	eqoał	zaki	daliŋ
WOMAN	mace	iʃa	ʕameni	tamətut	wahine	fefine	hɔni	vehivavi	pein	babae	nmatu	feto	perempuan	binanauʃað	tazuŋan	mqedin
MAN	namidʒi	iʃ	ħawata	ataras	kanaka	taŋata lahi	fa	lehilahi	tamol	lalake	kano			ajuði	ʁunanaj	rseno
WHITE	fari	lavan	ʕawak	aʃəmrar	keʔokeʔo	hinehina	fisi	fotsi	malkouk	puti	tar	mutin	putiħ	mapuði	busaʁ	bhege
BLACK	baki	ʃaxoʁ	boʕ	aβarʃan	ʔeleʔele	ʔuliʔuli	kele	mainti	tdomtbun	itum	got	metan	hitam	maqosum	tŋen	qalux
HOT	zaɸi	xam	daʕaʕam	aɣ	wela	vela	sunu	mahamai	wanana	init	ftin	manas	panas	mahnar	skwaru	mtilux
COLD	sanji	kaʁ	tsaʔ	asəmað	anu	momoko	matiti	maŋatsiaka	barum	bugnaw	mlanr	malirin	dingin	mahaðiwhiw	sen	msekuj
HERE		po	diri	ða	ʔaneʔi			eti	je	adia	sa	iha neʔe	di sini	inaj	zi	hini
THERE		ʃam	didaʔ	ðin	laila			eri	a	atua	san	iha neʔeba	di sana	isahaj	tajan	hija
LONG	t͡sawo	aʁok	tɬer	azijra	loa	loloa	roa	lava	mlae	tas	naframwen	naruk	panjaŋ	maqolijus	ʁruŋ	qnedis
SHORT	gad͡ʒere	katsar	kuqumar	aquðað	pokole	nounou	luka	fohi	katka	mubo	mit	badak	pendek	luiʃ	kzu	blebu
NIGHT	dare	lajla	xwera	d͡ʒirət	po	poʔuli	poŋi	alina	tdom	gabil	pog	kalan	malam	mahumhum	ʁabi	keman
DAY	rana	jom	bal	nha	la	ʔaho	terɔni	aŋdo	adjan	adlaw	mau naliati	loron	hari	qali	lan	dijan
FULL	cika	male	hats	ʃua	piha	fonu	hoi	feno	awan isai	punu	tisoksok			mabaʃbaʃ	bisuʁ	steŋe
EMPTY	bakome a ciki	ʁejk	kahar	xwa	haka	maha	tafa	foana	kao	hawaŋ	pal	kaʔare	kosoŋ	aðað		
NEW	sabo	xadaʃ	ʕaben	zðið	hou	fɔʔou	fɔʔou	vaovao	fou	bagu	foum		baru	fəqlu	tasu	bgurah
OLD	t͡soho	jaʃan	qaren	amɔqran	ʔaoʔo	motuʔa	mafua	antiŋa	wagama	dan	motu	katuas	tua	sasað	ʔzan	tsmutsats
ROUND	kewajaje	agol	gumbalalaħ	aquraj	poepoe	fuopotopoto	kɔlkɔlu	boribori	tlanti	bukud	manopnop			mabuðoq	turun	mtumun
FLAT	bai daja		peħ	wata	palahalaha	lafalafa	teaptepa	lemaka	labaka	lapad	matit		litʃin	mabareð	tpajas	qepi
DRY	busaʃe	javeʃ	kahar	azəɣ	maloʔo	momoa	mamasa	maina	gos	laja	gar	maran	keriŋ	makutbað	iseŋ	mdeŋu
WET	d͡ʒ̃ikake	ʁatov	naʕ	uf	maʔu	hauhau	matmata	lena	gan	basa	sumsmo		basah	matuðu	zizi	mhuriq
WIDE/BROAD	faði	ʁaxav	intɬaxw	miriw	laula	laulahi	tafa	malalaka	baŋan tbun	lawaŋ	polplo		lebar	mabaram	tabaj	glahaŋ
NARROW	maras faði	tsar	iiraakw	ħsa	laʔiki	fasiʔi	ʔele	eti	pŋpoŋa	higpit		klot	sempit	maqtit	niku	dgehin
THICK	kauri	ave	diqi ur	gða	manoanoa	matolu	mafolu	matevina	mtnol	baga	matol	mahar	tebal	makuʃtor	ʁutuz	qtehul
THIN	siriri	ʁaze	diqi nina	azðað	wiwi	pahauhau	mahini	manifi	mrjasa	nipis	mrara	krekas	kurus	lagqesusaj	impis	rqeliŋ
SMOOTH	sumul	xalak	motsotsoq	arəqaɣ	malino	hamolemole	marmarɔri	malama	smut	hamis	mal	kaber	halus	madeqro	laziu	mtbale
ROUGH	kauʃi	mxuspis	xaslaslaʔ	aħaʃaw	hoʔolua	petepete	vɔnu	marao	raf	gansal	liuliu				nisan	qtehul
HEAVY	nauji	kaved	iloʔ	ðqər	kaumaha	mamafa	maha	mavesaŋa	mrwa	bugat	mten	todan	berat	mabrek	zineq	tshedin
LIGHT (not HEAVY)	maras nauji	kal	inslaħ	fsəs	mama	maʔamaʔa	tʃeamtʃema	maivana	pasama	gan	mram	kman	riŋan		ineŋat	tslokah
DARK	duhu	xaʃuk		ħðəq	makuʔe	fakapoʔupoʔuli	maksulu	antiŋa		sulup	nmalko		malam	mahumhum	kumit	kuŋ
LIGHT (not DARK)	haske	vahiʁ		aʃəmrar	halakea	mama	maʔa	tanora		hajag	sor				tum	ledax
FAST	sauri	mahiʁ	ganslaj	s tazra	ʔawiwi	vave	væve	haiŋgana	wlwalemi	kusaj		lalais	ʧepat	mabiskaw	kramkamut	knuwa
SLOW	maras sauri	iti	tsegis	ʃwaj ʃwaj	lohi	tuai	fepi	miadana	pasak nami	hinaj	mailum	neineik	lamban	miaqawan	luŋut	knhuwe
HARD	tauri	xazek	gawid	qsəħ	ʔoʔoleʔa	fefeka	moumou	mafi	sakar	tiga	kerkerai			makuʃrak	qatmu	sadux
SOFT	lauʃi	ʁak	wananaʔ	aðəβ	palupalu	molu	maru	malemi	matala	humuk	mailum		lembut	mimbulnu	puʁu	mhenuk
DEEP	zurfi	amuk	tsalʔi	ad͡ʒəɣ	hohonu	loloto	lala	lalina	kror	lalum		as	tiŋgi	marukruk		truma
SHALLOW	maras zurfi	ʁadud	parampeħ	aʕra	papaʔu	mamaha	ʔelʔele	marivo	ftfota	mabaw			daŋkal	makapa	itaza	
HIGH	bisa	gavua	gawa	aʕra	kiʔekiʔe	maʔoluŋa	lamlama	amboni	lak na	habug	elag	as	tiŋgi		ibabaw	baro
LOW	fasa	gamuk	baraj	aða	lalo	maʔulalo	ʔe	ambani	tan na	bahadu		badak	pende?	luiʃ		blebu
MOTHER	uwa	ima	ajo	jəma	makuahine	faʔe	oʔhoni	reni	tna	nanaj	raiten	inan	ibu	ina	na	bubu
FATHER	uba	av	baba	βaβa	makua kane	tamai	øʔfa	rai	tama	itaj	tata	aman	bapaʔ	ama	tama	tama

	Indo-European					Niger-Congo											
	Czech	**German**	**Breton**	**Armenian**	**Persian**	**Swahili**	**Sesotho**	**Zulu**	**Nyanja**	**C'Lela**	**Akan**	**Ewe**	**Kinyarwanda**	**Yoruba**	**Bambara**	**Wolof**	**Mandinka**
I	ja	iç	me	jes	mæn	mimi	na	mina		am	me	mɛ	ɲje	emi	ne	man	nte
YOU	ti	du	te	du	to	wɛwɛ	wena	wena	iwe	dan	wo	wo	mwebwe		i	i	ite
BIG	veliki	gʁos	bʁaz	mets	bozorg	kubwa	tse xolo	kulu	kulu	damra	kɛse	ga	kinini	tobi	belebele	rej	ba
SMALL	mali	klain	bihan	pokər	xord	dɔgɔ	tse ɲɛnane	nlane	ŋono	æsɔ	ketekete		kaninira	kere	dɔgɔ	tuti	doja
GOOD	dobri	gut	mad	lav	xub	ɛma	tse ntle	le	bwino	kemadpaska	pa	ɲo	itʃiza	idara	ɲuman	nɛx	beto
BAD	ʃpatni	ʃlext	fal	vat	bæd	ɔvu	tse mpe	bi	ipa	vdumu	bone	vloe	itʃiki	oda	go	bɔn	ɟawo
THIS	toto	disəʁ		uə		huju	ona	lo	uju	sin	ji		ikiŋiki	eleji	nin	ɟi	ɲiŋ
THAT	tamto	jenəʁ				julɛ	jane	lo	ujo	kav	no		itʃo	ejini	o	lɔlu	wo
MUCH/MANY	mnoho	fil	kalz	ʃat	zijad	iŋgi	tse ŋata	niŋi	mbiri	hwedi	baha	gbɔsɔsɔ	tkwinʃi	pupo	tɛa	fɔ	sija
FEW	malo	veniç	nebød	kitʃ	ændæk	tʃʃatʃɛ	malwa	nlane	ena	tʃikm	kakra	sue		die	dænin		domandiŋ
BEFORE	pred	foʁ	aʁaog	aradʒ	piʃ	kabla	pele	pambili		dam	ansana	ŋgɔgbe	mbere	siwadzu	jani		doto
AFTER	po	nax	gude	heto	pæs	bada	hoba	emuva	pambujo		tɛiri	megbe	haɲuma		kəfɛ		kola
ABOVE	nahore	ybəʁ	auz	verevə	bala	dʒu	hodimo	pezulu	mwamba	atʃona	so		haruguru	loke	sanfɛ	tim	oba
BELOW	dolu	untəʁ	en tʁaoɲ	nerkevə	paʔin	tʃini	tlasa	pansi	pansi pa	dbæbo	se		munsi	isale	dzukɔrɔ	rɔn	koto
FAR	daleko	feʁn	pel	heru	dur	mbali	hole	kude	kutali	naka	nihɔ	tsoabo	kure	ɟino	dzan	sɔrɛ	ɟanfa
NEAR	blizko	nahə	tost	mot	næzdik	karibu	pela	seduze	kala pafupi	dotræe	ɛbɛn	zɔ de dzi	hafi	itosi	kɛrɛ fɛ	ɟɛgɛ	bala
WOMAN	ʒena	fʁau	mauez	kin	zæn	mwanamkɛ	mosadi	umfazi	mkazi	næta	katasia	srɔɲɔnu	umugore	obirin	muso	ɟigen	muso
MAN	muʒ	man	gwaz	təɣamard	mærd	mwanamumɛ	mona	indoda	mwamuna	armɨ	bani	ɲutsu	umugabo	okurin	ma		ke
WHITE	bili	vais	gwen	spitak	sæfid	ɛupɛ	tse tsweu	mɬope	jera	ipusi	sako	ɣi	umweru	funfun	jɛ	wɛx	koj
BLACK	tʃerni	ʃvaʁts	du	sev	sijah	ɛusi	tse ntso	mɲama	da	tʃip	tuntum	jibɔ	itʃirabura	du	fin	ɲul	fiŋ
HOT	fiorki	hais	tom	ʃat tak	daq	jɔtɔ	tʃesa	ʃisa	mva kutenta	tʃudtʃudu	çeçew	xɔ dzo	igiʃjuʃje	gbona	gan	tanga	kandi
COLD	studeni	kalt	jen	sarə	særd	baridi	serame	makaza	zizira	tup	win	fa	igikonɟe	otutu	mura	sɛda	sumajata
HERE	tadi	hiʁ				hapa	teŋ	lapa	pano	taha	ha	afi	ahaŋaha	ibiji	jan	fi	ɟaŋ
THERE	tam	da				kulɛ	mo	lapo	apo	inlu	hɔ		harija	ibejen	jen	falɛ	woto
LONG	dlouhi	laŋ	hiʁ	jerkar	derɑz	rɛfu	tse telele	de	tali	pɔrɔrɔ	tenten		urudatsa	gun	dzan	guda	ɟanfa
SHORT	kratki	kuʁts	beʁ	kartʃ	kutaɦqæd	fupi	tse xutswaŋane	fiʃa	fupi	bitirma	tiawa		kigufi	kuru	surun	gata	sutija
NIGHT	nots	naxt	noz	giʃer	ʃæb	usiku	bosiu	ubusuku	usiku	gjopo	anadʒo			ale	su	gudi	suto
DAY	den	tak	deiz	or	ruz	mtʃana	moła	ilaŋa	tsiku	dhɔn	da	ŋkeke	umunsi	oɟo	don	bes	tilo
FULL	plni	fol	leun	li	por	dʒa	tletse	giwele	dzadza	rompim	petə	jɔ	itʃuzuje	kun	fa	fɛs	fa
EMPTY	prazdni	leʁ	gulo	datark	xali	tupu	fela	ze	pwa	pjangas	sa	ɸuɸlu	ubusa		kolon	nɛn	kenseŋo
NEW	novi	nɔy	nevez	nor	dʒædid	pja	tse ntjha	ʃa	tsopano	poja	kaseɛ		amakuru	tutun	kura	ɛs	kutaja
OLD	stari	alt	koz	tser	pir	mzɛ	holo	dala	kalamba	utu	dadaw		igiʃaɟe	pe	kɔrɔ	mag	kebajata
ROUND	okrouhli	ʁunt	ʁond	kəlor	gerd	duara	tse tʃitʃa	jindiliŋa	zuŋulira	dogrot	purutwa		gutsirita	ajika	kori	morog	mulumulu
FLAT	plotší	flax	plad	tapak	hæmvar	bapa	separa	labazeka	onoka	fjak fjako	tratra		uburambuke			taca	dampataŋ
DRY	sutší	tʁokən	seh	tʃor	xoʃk	kauka		omile	uma	kangram	wese	ɸuɸu	itʃumɲe	gbe	dza	wɔw	ja
WET	mokri	nas	gleb	tats	tær	lɔwana	metsi	manzi		tʃes	fɔtɛe		umuzulu	somi	ɲigin	tɔj	monto
WIDE/BROAD	ʃiroki	bʁait	ledan	lajn	pæhn	pana	batsi	banzi	takata	jatata	tetretɛ	gbadza	itʃigazi	gbengbe	dzosurun		fanu
NARROW	uzki	ʃmal	stʁiz	neɣ	tæŋg	ɛmbamba	tshesane		mpanipani	pɔrɔrɔ	hea		ihaŋa	hiha	dɔgɔman		dete
THICK	tlusti	dik	teo	hast	koloft	nɛnɛ	tse tena	nołonze	katamira	dik	mu duro		igitsindagije	ikpodzu	fire	far	kulija
THIN	tenki	dyn	tano	barak	laqer	ɛmbamba	tse tsesane	ondile	onda	felfele	tea		kinanutse	tirin	fasa	lapa	fema
SMOOTH	filadki	glat	fluʁ	hart	hæmvar	laini	boreledi	buʃelezi	salala	perpere	mbrɛw		itʃiɲerera	kuna	nugu		tembe
ROUGH	drsni	rauh	ʁust	koʃt	doroʃt	kwaruza			njambalaza	hu	wereɛwerɛ	nusi mesɔ o		hanahana			sakali
HEAVY	teʒki	ʃvəʁ	ponɛʁ	tsanər	sæŋgin	zito	boima	nzima	lemera	kniki		kpekpeme		wuwo	girin	dis	kuliŋo
LIGHT (not HEAVY)	lefiki	laixt	skaŋv	tetev	sæbok	ɛpesi	kxaŋa	lula		tʃdapku	heran				fɛgɛn	lɛr	
DARK	temni	dunkəl	teŋval	mug	tarik	giza	tse ntso	łwile	kwetʃera	tʃip	kusu	do viviti	umwijima		finman	ləndəm	dibiriŋ
LIGHT (not DARK)	svetli	hel	sklaeʁ	paytsar	rowʃæn	mwaŋga	bobebe	ukukaŋa	tuwira	mjaz	ani dum				kin		
FAST	rišli	ʃnel	buan	arag	zud	upesi	pakisaŋ	ʃeʃa	fulumiza	pirit	hare	tsi nu dɔ	kwigomɲa	deno	teliman		tariŋo
SLOW	zdlouhavi	laŋgsam	goʁeg	dandaɣ	jævaʃ	pɔlɛpɔlɛ	butle	donda	paŋono	kak	brɛw		buhoro		suma	jix	hakilikuma
HARD	tvrdi	hart	kaled	pind	seft	gumu	tata	lukuni		hoku	bene	sese	gukomera		gɛlɛnman	naxari	ɟa
SOFT	meki	wiç	dus	papuk	nærm	ɔrɔrɔ	bonolo	ntofontofo	fewa	bodbodo	tofotofo		itʃorosje	ero	magan		fojariŋo
DEEP	filuboki	tif	don	xor	ʒærf	rɛfu	tebileŋ	julile	kuja	pimka	dɔ	goglo			dun		dinoŋ
SHALLOW	melki	siçt	baz	tsantsaɣ		kina kifupi	se tebaŋ		sazama	dtisi							
HIGH	visoki	hox	uhel	bardzər	bolænd	rɛfu	pahameŋ	de	mwamba	ja	wɔ soro	kɔkɔ	hɛjuru		dzamandzan		ɟanajariŋ
LOW	nizki	nidriç	izel	tsats	pæst	tʃini tʃini		fiʃa		tedep	fam		hasi	dzidzin			
MOTHER	matka	mutəʁ	mam	majr	madær	mama	me	umama	maji	mama	na	nɔ	ɲina	ija	ma	ndɛj	ba
FATHER	otets	fatəʁ	tad	hajr	pedær	baba	ntate	ubaba	tate	baba	papa	tɔ		baba	fa	baj	baba

	Sino-Tibetan					Trans-New Guinea	Australian		Austro-Asiatic		Nilo-Saharan		Oto-Manguean	
	Mandarin	**Manange**	**Tibetan**	**Mikir**	**Akha**	**Fataluku**	**Gurindji**	**Warlpiri**	**Vietnamese**	**Cheq Wong**	**Ghulfan**	**Kanuri**	**Otomi**	**Chatino**
I	wo	ŋʌ	ŋa	ne	ŋa	ana	ŋaju	ŋaculu	tau	ʔiŋ	jɛ	wu	nuga	naʔ
YOU	ni	kjʌ	kjod	naŋ	nɔ	a	ŋuntu	ŋuntu	mai	miʔ	aj	ŋi	nuʔaʔ	nuʔi
BIG	da	tʃʌ	tɛɛn po	te	hy	lafai	caŋkaḷi	pumpukarimi	tɔ	mniʔ	adʒe	kura	data	tlu
SMALL	ɕiɑu	tʃa	tɕuŋ tɕuŋ	so	ti	kaʔukisa	japawuru	wita	nə	hiʔɔj	watunu	kabəgə	tolo	luʔ
GOOD	xau	sʌ	jag po	mesen	my	rau	puŋu	markuɾa	tot	bajek	kɛŋ	kədʒi	za	suʔe
BAD	xuai	asʌ	sdug tɕags	hiŋno	dø	mahune	wankac	juju	sau	həʔ majek	biɾu	kutu	hingar za	kuʃi
THIS	ʤɻege	tsu	di	la	hə	e	ŋawa	ŋampu	nai	daʔ		adə	nuna	
THAT	nage	u	pa gi	hala	tə	i	ŋila	kuja	dɔ	nuʔ		tudu	nuʔa	
MUCH/MANY	ɕuduɔ	peʔ	maŋ po	oŋ	mja	lafane	carwa	cajacari	ŋiəu	bsoʔ	əndʒiri	ngəwu	ʃi	kaʔa
FEW	xenʂɑu	njeta	ŋuŋ ŋuŋ	penaŋbak	tjy	jewaje	murkunpuru	maŋkurpa	it	kanɛt	watini	gana	tʃitho	tʃiʔ ti
BEFORE	itɕiɛn	naraŋ	ʂŋon la		mɛ hu	tara		muŋma	truək		or	tʃa	ʔbiʔbɛtho	lo
AFTER	ixɔu	tʃu tʃu	rjɛs la	dzutleta	mɛnaŋ	ta	ŋiŋi	ŋaka	ʂau	lpas	kal	ngawo ben	gemʔbu	tʃuʔ
ABOVE	ʂaŋmiɛn	piri	sgaŋla		ta	puhu	kankulapal	pajapaja	ʧen	hantən	twaj	samilan	maʃətse	lo
BELOW	ɕiamiɛn	naŋri	mar	aber	la o	ehitʃine	kaniɲcal		ziəi	kjɔm	ʧwale	tʃidija	ngati	niʔ
FAR	jyɛn	ta ruŋ	tag riŋ po	helokon	maŋ	tʃo	jikili	munpara	sa	tson	durdʒir	tʃində	jaʔbi	tihuʔ
NEAR	dzin	nepʌ	ɲɛ po	tebok	dɔpɛ	karu	kacupari	carara	gan	dwah	ɔkur	karən	getiʔbi	kiʔ
WOMAN	nyʐən	mo	skjɛs dman	arloso	za mi abu	tupuru	canka	kaŋta	dan ba	koŋ	ədu	kamu	ʔbɛhŋa	neʔ
MAN	nanʐən	po		pinso	xa dze za	maro	ŋumpit	wati	dan oŋ	tuŋkal	kortu	kwa	ʔɲəhə	ju
WHITE	bai	tʌrkja	dkar po	lok	pju	piti	waki	kaɾiri	ʧaŋ	puteh	oɾi	bul	taʃi	ngate
BLACK	xei	mleŋkja	nag po	ik	na	lakuvare	mumpuŋ	kiɾiʎkiɾiʎpa	den	səʔɛŋ	uri	sələm	ʔbo	ngata
HOT	ʐə	tse	tsa po	karom	tsa	timine	tupurŋ	karkarjankami	nɔŋ	bit	wi	kanua	ʃi mpa	tikeʔ
COLD	ləŋ	kʌŋ	graŋ mo	tɕuŋ	ga	ikare	makuru	kaʎkarimi	laŋ	sdɛts	urguŋ	kaku	ʃi ntsɛ	tlaʔ
HERE	ʤʐəʐ	tsuri	dir	dak	hə ga	e naʔe	muḷaŋka	ŋampu	dai	handaʔ		na adə	nuwa	
THERE	naʐ	uri	pa gir	hali	tə ga	ivi naʔe	jalaŋka	palka	kia	hantoj		na tudu	nuni	
LONG	ʧaŋ	ruŋ	riŋ po	kiŋ	maŋ	veluvelu	cuwal	kawaʈa	zai	tsiŋ	ɖodʒi	datəa	ʃi maki	tuki
SHORT	duan	tuŋpʌ	tuŋ tuŋ	tihek	njm	kava	lutu	micipuru	ŋan	tsinhuɛt	ʃɛrdu	kori	hingi ma	luʔ
NIGHT	jə	muntse	mtsan	dzirlo	u tji	kounu	wuḷŋaŋ	muŋa	dem	btom	ətɛri	bəne	nʃui	tla
DAY	tiɛn	sa	ɲi ma	nerlo	naŋ	vatʃu	kaputa	ŋulcurpa	ŋai	ktɔʔ	ulal	jim	mpa	ʃa
FULL	man	naŋ	kɛŋs	pleŋ	bjaŋ		timpak	cakaŋaʎa	dai	bək	əŋat	zəmbəli	ʃi ŋutsi	na
EMPTY	koŋ	te	stoŋ pa	kedzoi	ŋɛ		taŋka	ɾiʎpirkarimi	ʧoŋ		kominŋ	de	hinti o	laha
NEW	ɕin	tsaŋ	gsar pa	akimi	sjy	miri	calajalaŋ	mintilpari	məi	rɛʔ	er	bəlin	ʔraʔjo	ki
OLD	lau	kjokor	ɾniŋ pa	sar	ø	kuare	kaciri	kumalpa	gia	badɔn	torɾa	tʃari	ʃita	kula
ROUND	juan	kikil	sgor sgor	boŋloŋ	laŋ		piṭakpari	cakuɾanpari	ʧɔn	twaw	ʧur	dukurkur	tsanti	ski
FLAT	piŋ	taŋ	lɛb lɛb	klam	ʃa	hila	palki	laʎpa	baŋ		tordʒa	fele	nidi	lka
DRY	gan	karkjʌ	skam po	kreŋ	gy	titile	pulwar	lalka	xo	kriŋ	ʃwat	ngamdə	ʃi ʔjoti	wti
WET	ʂi	ʃu tʌ	rlon pa	tɕam	a	tʃulu	pantic	miɲminpa	uɔt	məʔɑc	orgi	kəli	ʃi mobo	lʃaʔ
WIDE/BROAD	kuan	tuŋpʌ	gu jaŋs po	dzam	dje	maluere	pajal	caḷawanawana	ʐoŋ	mniʔ	adʒe	faran	ʃi ʃogi	ʃe lo
NARROW	ʤai	kje to	gu dog po	tɕeŋ	su	malete	kurpu	pururu	hɛp	gats	watunu	zəka	ʃi ntsiʃti	lti ti
THICK	xou	ʂuŋ	mtug po	iŋtɛo	tu	rapake	cuɲcu	lumuɾucarimi	dai	təl	koloka	gadʒadʒak	ʃi mpidi	tnu
THIN	bo	pɾʌ	srab po	tɕuŋkreŋ	dje	lika	tinkal	larpalarpa	məŋ	stal	oʈur	tənene	ʃi ntsi	ti
SMOOTH	xua	tʃa	dzam po	nei	djy	hilate	kiriŋkiriŋ	kancaʎkancaʎpa	ŋan	sluts	ɖuri	sulsul	ʃi ntsiki	kuti
ROUGH	tsutsao		rtsub po	puroiprok	sa		puṭuwuṭu	ŋinṭiniṇti	ʐap	kasar	koɲi	kərsasa	ʃi nzadi	ʧe?
HEAVY	ʤʐoŋ	li	ldzid po	ardik	kaŋ	tʃiʔire	ŋaṭur	cantakupari	naŋ	lbat	ʧilur	kurwowu	ʃi hɲi	tiʔi
LIGHT (not HEAVY)	tɕiŋ	ŋaŋ	jaŋ po	iŋar	pja	vakahe	rampali	rampaku	ŋɛ	hajoŋ	ʧilinɖu	kamboì	ʃitho	sa
DARK	an	mona	nag kuŋ	iŋtiŋ ik	zaŋ	koukoune	mum	maru	ʂam	skuj	uri	nəmkəlməske	mbotsi	tla
LIGHT (not DARK)	liaŋ		od	karaŋ	bja		ŋunti	cara	ŋat	puteh	oɾi	nura	ʃi nzi	luwi
FAST	kuai	kini	mgjogs po		kɔ	nahake	wacica	kapanku	ŋaŋ	tspat	ŋadʒuŋ	nagat	nsəni	ndla ti
SLOW	man	kole	ga le	iŋlen	maŋ	teile	jamak	kuɻuḷu	cam	dmij	ŋaŋindu	jilajen	ŋante	tija?
HARD	jiŋ	kjoŋ	mkrɛgs po	iŋtaŋ	ɣaŋ		patawaŋ	cumati	cuŋ	tsgih	bur	ʧibu	ʃi me	tla
SOFT	ʐuan	ʧe	sob sob	iŋduk	naŋ	romo	kaljpa	kaʎirpa	mem	lmən	oti	təlala	ʃi ntudi	hoʔo
DEEP	ʂən	tiŋ ara	gtiŋ zab po		na	lohai	taḷukuru	ɾaku	sau	dalam	ɖodʒi	kuruwu	ʃi hɲei	kiʔɲi
SHALLOW	tɕian	mʌ		iŋdei	ma na	kava	kalca	calca	noŋ		ʃɛrdu	dangərese	hingi hei	lti ti
HIGH	gau	no	mto po	iŋtui	go	lohai	ŋaramala	kankaḷu	kau	tiŋgiʔ	ʧwaj	datəa	ʃi hɲetsi	ka
LOW	ai	naŋ	dma po	iŋdei	ɔ	kava	kaɲcal	waʎapiṇṭipiṇti	tap	katɛʔ	ʧwale	ʧidijan	meʃtha	ŋna
MOTHER	mama	amʌ	a ma	pei	ama	nalu	ŋamaji	capunku	mɛ	məj	iniŋ	ja	nana	haʔa
FATHER	baba	apʌ	pa pa	po	ada	palu	ŋaci	cacipaɾu	bo	bap	agaŋ	bawa	tada	sti

	<7 languages per family			7-20 languages per family			21-44 languages per family					45-68 languages per family				
	Ayoreo	**Basque**	**Korean**	**Epena**	**Karok**	**Scando-Romani**	**Yaminahua**	**Estonian**	**Tlingit**	**White Hmong**	**Nama**	**Aché**	**Nahuatl**	**Wapishana**	**Imbabura Quechua**	**Turkish**
I	ju	ni	na	mɨ	na	mander	ɨ	mina	xʌt	kuv		tʃo	nehwa	ugarɨ	ɲuka	ben
YOU	ua	hi	nu	pɨ	ʔim	diru	mi	sina	wʌe	koc		ʤe	teh	pigarɨ	kan	sen
BIG	kerui	lodi	kɯn	netʃoma	ketʃ	baru	ɨβapa	sur		loc	kai	watʃu	weji	ɨdarɨ	tʃatun	byjyk
SMALL	atiami	tipi	jak ɯn	kaipe	ninamitʃ	lala	piʃta	wæike	gek	me	ǀa	pɨkae	ʃikʃin	kaikapasudiʔu	utʃiʎa	kytʃyk
GOOD	arei	hun	tɕo ɯn	pia	jaβ	kisku	ʃara	hea	ke	zo	!ai	gatu	kali	kaiman	ali	iji
BAD	sinai	gaiʃto	na bun	katʃia	kem	noɕlu	tʃaka	halb	ʃike	pem	tsu	butʃa	amokali	idikaida	miʎaj	hoʃ olmajan
THIS		hau				kava		se		no	ne				kaj	bu
THAT		hura				dole		se		ntawd	ǁa				tʃaj	ʃu
MUCH/MANY	uʔasena	hanitʃ	ma nun	tʃokara	taj	but	itʃapa	palju		cob	ǂui	tara	katʃi	iriba	atʃka	tʃok
FEW	garosi	guti	tɕo gum	makɨaraʔe	tʃimitʃ		piʃta	wæhe		mentsis	ǀoro	prowi	katʃi tipiʃin	maskaidaʔɨ	aʃa	birkatʃ
BEFORE	iʔkaiteʔ	lehen	tɕun eh	na		gjal	βisona	ene		tom ntec		tʃa ruwa	atʃto	ɨa i	ɲaupa	bir ʃejden
AFTER	digaʔ	gero	ho eh	te		akter	tʃipo	pærast	a it	tom qab		bu		dauwa ati	tʃipa	sonra
ABOVE	gate	gaɲen	we eh	irɨ	ʔaʔ	œfters	βomakia	ylewal	a kinak	saum	ǂora	jabu	ahko	duku i	tʃawa	jukaruja
BELOW	ʔudi	pean	mi te	edu	suruk	telal	maikiri	al	a seji	hauv qab		widʒi	tani	warani i	uraj	aʃaɯda
FAR	idahaʔ	hurun	mun	timia	jiβ	duri	tʃai	kaugel	le	deb	!aru	budʒaja	wehka	minapu	karu	uzak
NEAR	idogosiʔ	hulan		kaita	ʔumukitʃ	niga	tʃaimaʃta	læhedal	du xʌn	ze	tawa	ripija	amo wehka	maunapa	kaiʎaʎa	jakɯn
WOMAN	ajore	emaste	ju tɕa	wera	ʔasiktaβan	maj	kiro	naine	ʃawʌt	pocniam	kxoes	kudʒa	siwat	zina	warmi	kadɯn
MAN	ajorei	gison	nam tɕa	imikira	ʔaβansa	gae	noko βinɨ	mes	ka	txiv nec	kxoeb	kbaeʔe	tagat	daunaiura	kari	adam
WHITE	pororoi	ʃuri	ha jaɲ	toro	tahko	parno	oʃo	walge	tlet	dawb	!uri	idʒu	istak	barakaʔu	juratʃ	bejaz
BLACK	utatai	belɕ	gum tɕuŋ	paima	ʔikxaramkunitʃ	kalo	βiso	must	tutʃ	dub	ǂnu	braʔa	tiltik	pudidiʔu	jana	kara
HOT	notari	bero	duɯ gu on	wasia	ʔimfir	bakalo	ʃana	kum	jʌjʌta	kub	ǁχoasa	aku	totonik	witʃaʔu	kunuk	suɯdʒak
COLD	tenui	hoɕ	tɕan	kurasa	ʔaθik	tʃylmert	maɕi	kylm	at	txias	!χai	duiʔɨ	sesek	wadidiʔu	tʃiri	souk
HERE		hemen	ju gi			kaj		sin		qov no					kajpi	burasɯ
THERE		han	tɕu gi eh			dik		seal		ntawd					tʃajpi	var
LONG	joʔikoi	luse	gin	teso	varam	dotʃu	tʃainipa	pik		ntev	gaixu	puku	wejak	dawɨ	suni	uzun
SHORT	katadei	labur	tɕal bun	kakatua	ʔipʃunkinatʃ	tiknu	tiʃopiʃta	lyhike	jatɬ	luv		kita	ʃikʃin	madiwai	utʃiʎa	kɯsa
NIGHT	dehai	gai	bam	parikua	ʔikxaram	rati	βakiʃ	ø	tat	tsaus ntuc	tsuxub	tʃawa	jowal	wamariɬan	tuta	gedʒe
DAY	diri	egun	nat	ewari	supah	devis	pina	pæew	jʌkji	hnub	tse	kreibu	tonal	wakandan	puntʃa	gyndyz
FULL	iragui	beterik	gwak tɕan	ipuru	ʔaxjar	perdu		tæis	hik	puv	ǀoa	ebema	peʃontok	paida	tʃunda	dolu
EMPTY	ahegesai	huɕik	bi uh it nun	arahaga	ʔarun		ʃaka	tyhi		kob		ija	kinokia	mawazu	tʃuʃatʃ	boʃ
NEW	itʃai	beri	sɛ lo on	tʃiwidi	pit	nevu	βina	us		tsiab	ǀasa	kreware	yankwik	paʔinaʔu	muʃutʃ	jeni
OLD	tʃokinai	sahar	oh lɛ doin	tʃisore	keβrikha	puru	ʃini	wana	ʃan	qub	ǀoro	tʃueʔe	soltik	zam	ruku	jaʃlu
ROUND	katadei	biribil		porokoko	ʔuruh		tokoro	ymargune		keckec	ǂuwu	papi	jowaltik	kadazadaʔiʔu	mujundi	juvarlak
FLAT	ehai	sabal	pjuŋ pjuŋ han	heweda	tinihitʃ		sapa	lame	tal	pliab	ǂa	tʃepe		madaz	pambaʎa	daire
DRY	kanori	idor	gun tɕo han	puasa	ʔiβaxrah	tostu	mito	kuiw	jaxuk	quav	ǂna	ʤɨpi	kiwaʃa	aradaʔu	tʃakiʃka	kuru
WET	tetai	umi	tɕu tɕun	beke	ʔaskipatʃ	kindu	mitʃa	mærg	jux	ntub	ǀa	piri	paltik	zabiʔu	ʃutuʃka	ɯslak
WIDE/BROAD	dakodei	sabal	nul bun	hobia	tirih		aʃβa	lai	wux	dav	hara	kira	patawak	kazu	antʃu	geniʃ
NARROW	ahami	hersi	tɕo bun	pirupe	tʃujitʃ		aɕo	kitsas	sa	nqaim	ǂo	mini	kopiʃtik	mazu	kitʃitʃi	dar
THICK	daɲorei	lodi	do gu on	pada	ʔitpum	groɕno	nami	paks	tla	tuab	!au	kira	tilawak	midi	raku	kalɯn
THIN	godoi	mehe	jal bun	namia	xutnahitʃ		βiʃta	penike	lixun	ɲias	ǁa	mini	kakaltik	madidika	ɲaɲu	indʒe
SMOOTH	karunai	leɲ	bo duɯ lu on	kakaja	xus		jiβa	sile	jas	du	ǂau	pretʃa	alaʃtik	miduda ʔu	amukʎa	dyz
ROUGH	tʃo tʃiki tʃĩki	laɕ	gu tɕin	ikia	kukuh		βiʃki	kare	kasixax		ǀχora	kuru	ʃihʃipintik	kaduraizabaʔu	ʃagra	pyryzly
HEAVY	penui	pesu	mo gu on	tʃia	maθ	drumit	iβi	raske	jʌdal	hɲav	!om	puiʔɨ	etik	kimiʃaʔu	ʎaʃatʃ	aɯr
LIGHT (not HEAVY)	punui	ahin		wesa	ʔiʃnanitʃ		ʃaka	kerge		sib	suwu	puija	ehkawtik	mamiʃaʔu	fangaʎa	
DARK	topiʔei	ulhun	uh do on	poʔre		tamlo	βiso	tume		tsaus		pira		pudidiʔu ɨtan	janaʎa	karanlɯk
LIGHT (not DARK)	kirike	argi	bal gum	porepore		valgurna	oʃo	hele		kac	!a	idʒu		iʃazabazaʔu	puntʃaʎa	ɯʃuk
FAST	ibisoi	laster	bal li	isapai	tʃemjatʃ		koʃi	kire		sai	!ae	bitaja	ihsiwik	kadimanaʔu	utja	
SLOW	oronai	baratʃ	tɕun tɕun hi	piaka	tʃakaʔimitʃ	durument	βinamaʃta	aeglane	tʃijak	mac mam	ǂau	takuaja	jolik	kunaʔi	alimanda	javaʃ
HARD	etoi	gogor	dak dak han	tʃare	sakriβ	horsnu	kiriʃ	kəwa		tawv	!ari	krere	takawak	dadara	sintʃi	sert
SOFT	paratai	mardo	bo duɯ lu on	poreke	xetʃitʃ	mjukra	patʃi	pehme	dixatʃ	muag	tsam	praru	jemaʃin	ʃumadaʔu	amukʎa	jumuʃak
DEEP	idaha aha gesi	barna	gip ɯn	napia	nukah	duriteli	noa	sygaw		tob	!am	puku	wehkatan	kana	tʃundu	derin
SHALLOW	gareroi	amiɲi bat barna	ja tun	itapa	ʔasaxnahitʃ		tʃaima miɕama	madal	kajaʃan	ntiav	ǂe	kraja		mana	aʃaʎa	su
HIGH	idaha ihi numi	gora	nop ɯn	iti	ʔaʔvarih		kija	kərge		siab	ǀawi	iwa	wehkapan	dukuʔu	tʃatun	jyksek
LOW	idogosi ihi numi	apal	na jun	hepeda			namaʃtai	madal		qis		kuru		apuwaʔa	utʃiʎa	altʃak
MOTHER	date	ama	uh mu ni	nawe	tat	dagri	iβa	ema	tla	niam	is	egi	inanʃin	daru	mama	ane
FATHER	daje	aita	ah bu tɕi	akore	ʔaka	dad	ipa	isa	iʃ	txiv	ib	apa	itahʃin	darɨ	tajta	baba

	68-92 languages per family		Replacement languages									
	Seychelles Creole	Thai	Catuquina	Ket	Japanese	Archi	Tok Pisin	Ainu	Inuktitut	Cheyenne	Zinacantán Tzotzil	Georgian
I	mon	tɕan	ia	at	wataɕi	zon	mi	kuani	uvaŋa	na		me
YOU	ou	kun	mia	u	anata	un	ju	aoka	ilisi	ne	voʔot	ʃen
BIG	gʀo	jaj	aɲipa	qa	oki	dozut	bikpela	poro	aɲijuq	haɦp	muk	didi
SMALL	pti	lek	ʃapitʃa	hɔna	tɕisai	titut	liklik	hutne	mikituq	tʃeʃke	bikit	patara
GOOD	bon	di	roʔapa	aqta	i	hibatu	gut	pirka	piujuq	epeva	lek	kargi
BAD	move	raj	itʃapa	sel	waɾui	χalut	nogut	wen	piunŋituq	havesev	tʃopol	tsudi
THIS	senla	ni		kire	koɾe	jamu	dispela	tan		heʔtohe	liʔ	es
THAT	sa	nan		qare	aɾe	tow	olsem	toan		haʔtohe	tah	is
MUCH/MANY	bokou	laj	otipa	on	oi	labɣanʃi	planti	porono	amisut	haesto	ʔep	bevri
FEW	pe	sɔŋsam	otimapitʃa	qomat	sukoɕi	tina	wan wan		pitsalaqpuq	tohɦkom	ɦaj	tsota
BEFORE	avan	kɔn	βiβo	qotils	mae	harak	bipo	hoski	sivuliuqpa		ta	adre
AFTER	apʁe	laŋ	ʃipo	qariɣa	ato	χita	bihain	okakeankor	kiɲulirmik			ʃemdeg
ABOVE	anlɛʁ	nua	manaʔori	ət	ue	jatik	antap		qulani	heʔam	akol	zevit
BELOW	anba	taj	wakaʔori	huɯtka	ɕita	kɬarak	ananit		kitani			kvevit
FAR	lwen	klaj	toki	bil	toi	aχʃi	loŋwe	hankeko	uŋasiktuq	ɦaʔeʃe	nom	ʃors
NEAR	koste	klaj	oraɯa	ɯlga	tɕikai	ɬak	klosap	sama ta	qanituq	kaɦkese	nopol	axlos
WOMAN	fanm	pujiŋ	aiβo	qim	dzosei	ɬonol	meri	menoko	arnaq	ɦeʔe	ʔantz	kali
MAN	zonm	putɕaj	honi	ket	dansei	boʃor	man	okajpo	aɲut	hetane	mol	katsi
WHITE	blan	kaw	hoʃo	taɣam	ɕiɾoi	tʃubatut	wait	retar	qauluqtuq	evoʔkomo	sak	tetri
BLACK	nwaʁ	dam	ʃiʃi	tum	kuɾoi	beχetut	blakpela	kune	qirniqtaq	moʔkohta	ʔik	ʃavi
HOT	so	rɔn	ʃana	aŋ	atsui	ʁilitut	hat	sirsesek	usuvituq	haohoʔta	kok	tbili
COLD	fʁe	naw	maɕi	qaqtuɯm	tsumetai	χetut	kol	merajke	ikiqpuq	etoneto	sik	tsivi
HERE	isi	tini		kiseŋ	koko	iʃik	hia	teta		heneheno	liʔ	ak
THERE	la	tinan		qaseŋ	asoko	imik	loŋ ɦap •	toanta		neheohe	tah	ik
LONG	lɔng	jaw	kijapa	ugde	nagai	aqatut	loŋ	tane	takijuq	ehaʔeso	natil	grdzeli
SHORT	kouʁt	san	tokopitʃa	hitim	midzikai	kutatut	sot	keweram	naituq	kaʔ	tsapo	mokle
NIGHT	lanwit	klaŋkun	jame	si	joɾu	iʃ	nait	antsikar	unuaq	taʔe	ʔakubal	ɣame
DAY	lizouʁ	klaŋwan	βari	i	çi	iq	de	tokaptsup	uluq	eʃe	kakal	dɣe
FULL	ʁanpli	tem	maʃa i	qo	mitɕiɳu	atsutut	pulap	sikno	tatatuq	oʔkotomoʔena	noh	savse
EMPTY	vid	waŋ	ʃaka	quj	kaɾa	atʃatutut	klia	siroha		ponohta	ʃokol	tsarieli
NEW	nouvo	maj	βinapa	ki	ataɾaɕi	matsatut	nupela	asir	pigivuq	mon	ʔatʃ	axali
OLD	vye	kaeae	ʃinira	sin	fuɾui	ɣalatu	lapun	husko	utuqaq	enaʔhe	kaʔ	dzveli
ROUND	ron	mon	toro	kruglaj	maɾui	gukitut	raun	sikanatki	amalutuq	aʔese	setset	mrgvali
FLAT	plat	baŋ	ɕapa	haksem	taiɾa	qatut	stret	upaksine	nalimaktuq	kaɦkese	patʃal	brtqeli
DRY	pa fon	haeaeŋ	toaʃ	toʁojiŋ	kawaku	quretut	drai	riwa	paniqtuq	oʔ	takin	mʃrali
WET	imid	piak	miʃa	ultu	tsumetai	tʃaratut	i gat wara	petʃi	kinipajuq	heʔkov	tuʃul	sveli
WIDE/BROAD	laʁz	kwaŋ	naɭβa	qil	hiɾoi	qatut	brait	para	siliktuq	ɦahpeʔ	ʃemel	parto
NARROW	etʁwat	kaeaep	naɭβama	toʁ	semai	qaqartu	i no brait	hutne	mikituq	tʃeʃkeʔeotse	ɦapaɦtik	vitsro
THICK	epe	na	kiβi	bol	atsui	ditʃatut	stroŋ	irone	ijujuq	haonov	pim	skeli
THIN	mens	baŋ	kimima	haksem	usui	kalatut	bun natiŋ	kapar	satuq	kaɦko	hoj	txeli
SMOOTH	maʁbʁe	lamun	tʃiro	ul	subesube	tʃula	stret	rarak	qairatuq	hesox	tʃulul	stsori
ROUGH	bʁit	jap	ʃato	qɔgbareŋ	aɾai	qarqatut			kigjaktuq	peʔpeʔ	tʃiʃaltik	uxeʃi
HEAVY	louʁ	nak	iwipa	sə	omoi	iʁdut	hevi	pase	uqumaituq	hanae	ʔol	mdzime
LIGHT (not HEAVY)	leze	bao	ʃatapitʃa	bejiŋ	kaɾui	salatut	i no hevi	ratsako		evehpanano	sipson	msubuki
DARK	sonm	mut	ʃiʃi	tum	kuɾai	matʃatut	tudak	ekurok	isutuq	aʔenoʔn	tupet	muki
LIGHT (not DARK)	klɛʁ	sawaŋ	koro	kən	akaɾui	iqdut	tulait	kosne		evoʔneto		nateli
FAST	oplivit	rew	raja	dɔqta	hajai	ɣabkul	kwik	nitan	tukakaju	neʃev	suhem	stsrapi
SLOW	lant	tɕa	tanapitʃa	unat	noɾoi	okur	slo		sukaituq	hosovahe	kun	neli
HARD	diʁ	kla	jowi	but	katai	tanqdut	hat	niste	tisijuq	heʔkon	tsots	magari
SOFT	dou	num	waʃo	homilaŋ	jawaɾakai	qanatut	malmalum	hapur	aqituq	heʔke	kun	rbili
DEEP	pʁofon	luk	oki	hoʁ	fukai	jakdut	godaun tru		itijuq		natil	ɣrma
SHALLOW	pa fon	tun	okima	tot	asai		i no daun	ohak	ikatuq	tʃehеʔkotame	hoj	tsqalmtsire
HIGH	o	suŋ	oki	tojga	takai	beχutut	antap	ri	puqtujuq	haʔehoʔoese	tojol	maɣali
LOW	ba	tam	okima	hitim	çikui	huliʃi	daun	ram	pukituq	tox	pekel	dabali
MOTHER	manman	maeae	iwa	am	haha	buwa	mama	hapo	anana	ʃke	meʔil	deda
FATHER	papa	pɔ	papa	op	tɕitɕi	abtu	papa	atsa	atata	heh	totil	mama

Appendix B: Unreliable Deviation

Deviation of phoneme distribution compared to average of unreliable sound groups, including Voiceless Palatal and ə-like, as well as all sounds groups of the Combination-grouping. 50 % overrepresentation (+ in light green), 100 % overrepresentation (++ in green), 50 % underrepresentation (- in light red), 75 % underrepresentation (-- in red).

	Voiceless Palatal	ə-like	Voiceless Palatal F	Voiceless Palatal P	Voiceless Alveolar F	Voiceless Alveolar P	Voiceless Velar F	Voiceless Velar P	Voiceless Labial F	Voiceless Labial P	Voiced Palatal A	Voiced Palatal F	Voiced Palatal P	Voiced Alveolar A	Voiced Alveolar F	Voiced Alveolar P	Voiced Velar A	Voiced Velar F	Voiced Velar P	Voiced Labial A	Voiced Labial F	Voiced Labial P
I				-					--	--	+	--	--	++	--		++	--			++	--
YOU	--		--	-			-		--	-	+	++	--	++	--	+	--	--	-			--
BIG	-	-	-				-					++	--	++	--	++		++				++
SMALL	-			--	++	++	-	+	--			--	--	++	--	+				-	--	
GOOD	-	-	-	-	+				--			++		++	--	++		--			++	++
BAD	-	-		--	++							--	++	++	--	+					++	
THIS	--		--	--	+				--	-	++	++	++	++	--	+			-	--		
THAT	--	-	--	--	+	+			--	--	+	--	++	++	--	+			+	-		--
MUCH/MANY		-			++				++			++	--	++	--	++						++
FEW	-	--		--	++	++						--	--	++	--	++		--				
BEFORE	-	-		--	+				++			++	--	++	--	++					++	++
AFTER	-	--		--	+				++			++	--	++	--	++	+		+			++
ABOVE				-	+				++			++	--	++	--	++		+			++	++
BELOW		-		-	++		-		--			++	--	++	--	++		+	-		++	++
FAR	-	-	--	-	++					-		++	++	++	--	++						
NEAR	-		-	-	+			+	++			++	+	++	--	++	--			-		
WOMAN	-	-	-		+		-		++			++		++	-	++			-		++	
MAN	-	-		--								++	--	++	--	++	+				+	
WHITE		--		-	++	+			++			++	++	++	--	++		++			+	++
BLACK	-	-		--	++	+						--	++	++	--	++	+	++			++	++
HOT	-	-		--	++	+						--	++	++	--	++	+	++			++	++
COLD		-		--	++	+	-			--		--		++	--	++	+					
HERE	--		--	--			+		++			++	++	++	--	++	++			--		
THERE	--	--	-	--			+					++	--	++	-	++		--		-	+	
LONG	-	--	-	-	+		-		++			++	++	++	--	++	++	++	+	-	++	--
SHORT	--	--	-	--	++	++			++			++	--	++	--	++		+				
NIGHT	--		--	--	++	+			--			++		++	--	++						++
DAY	-	--	-	-	+				--			--	++	++	--	++	+		-		++	
FULL				-	++	+			++	+		++		++	--	++	+					
EMPTY		-			++				++			++	++	++	-	+	++			-		
NEW	-	--		--	++		-		++			++		++	--	++					++	
OLD	-	--	--	-	++							--	++	++	--	++	-	--			+	
ROUND			--				-					--	+	++	-	++		+	+			
FLAT	-	--	-	-	+	+			++	++		++		++	--	++	++	--	--	-		++
DRY	-	-	-	--	++				+			++	--	++	--	++	+	+	+		+	
WET		-	+	-	++	++						++	--	++	--		--				+	
WIDE/BROAD	--	--	--	--	+				++			++		++	--	++	+	+			++	++
NARROW	-	--		--	++	++				+	--	++	++	++	--	++		++				
THICK		--				+	-		++			++		++	--	++					++	
THIN	--	-	-	--	++				++			++	--	++	--	++			-			
SMOOTH	-	--	--		++					-		++		++	--	++	--	++	-			
ROUGH	-	--		-	++							++		++	--	++	-			-		
HEAVY			--		+						-	++		++	--	++	+	++			++	
LIGHT (not HEAVY)	-	--	-	--	++				++			--	++	++	--		++				++	
DARK	--		-	--	++	+					-	--	++	++	--	++	+	--			++	
LIGHT (not DARK)	-	-	--		+				--			++	++	++	--	++	+	++			++	
FAST	-	--	--		++							++	--	++	--	++	+				++	
SLOW		--			++						+	--		++	--	+		+			++	
HARD	-	-	--		++	+			++	-		--		++	--	++			++		+	
SOFT	--	--	-	--	+				++			++		++	--	++			-			
DEEP	-	-	-	--					++			++	++	++	--	++		++	+	-	--	
SHALLOW		-			++	+			++			--	-	++	--	++		+			+	
HIGH	-	-	--	--								++	++	++	--	++	++		++		+	++
LOW	-	--	-		++				++			--	--	++	--	++		--	-			++
MOTHER	--	-	--	-				-		-		--	--	++	--	+			-	+		
FATHER		-				++		-	++	++		--			--	+			--	-	++	++