Semantic Relations in Diachronic Word Families

This paper investigates semantic relations in semantic change, word formation, and borrowing in the etymological histories of 480 German nouns of basic vocabulary. Contiguity is shown to be the most frequent semantic relation, followed by taxonomy, confirming previous research. The use of diachronic data and analysis by multiple factors however demonstrate that the choice of the kind of semantic relation in creating new lexical meanings is strongly dependent on various aspects. Those include the part of speech of the words involved, the semantic field of the concepts, and whether the semantic relation is found in normal semantic change, in semantic change accompanying borrowing from another language, or in word formation, in the last case further dependent on the kind of word formation. Additionally, significant differences are found between attested and reconstructed parts of etymologies, providing insights into possible points of improvement for semantic reconstruction.

Keywords: semantic change, word formation, lexical typology, semantic relations

1. Introduction

Words in a language’s lexicon are historically related to each other in various dimensions, among them the semantic and morphological dimensions (Koch 2001: 1156-1159). Semantic relations can be found 1) within one word, namely between the original meaning and new meaning of words undergoing semantic change and 2) between two words, if one of them is derived from the other word via a word formation process.

For example, the German word Glück ‘good fortune, happiness’ used to mean only ‘good fortune’, so there is a semantic relation to be found between the original and the new meaning in this process of semantic change. Back in Middle High German, when gelücke had yet still another meaning, ‘(good or bad) fortune’, ungelücke ‘bad fortune’ was derived from it. Therefore, also here exists a semantic relation, in this case between the source and target of this word formation process (Figure 1).

In this paper, I study these relations from an onomasiological perspective, investigating the history of words referring to a set of concepts and the semantic pathways chosen for designating them (Blank 2001: 7).

The data used are 480 German nouns and their etymological histories. The set of nouns was chosen based on the concept list of basic vocabulary by Dellert et al. (2020) by restricting their German word list to include only concepts expressed by nouns. The etymologies are based primarily on the information provided in Kroonen (2013) and Pfeifer et al. (1993).

Koch and Marzo (2007) use a similar approach by studying formal and semantic relations in the 500 most frequent lexical words in French. While their study is based on synchronic forms and meanings alone, I also take historical (attested or reconstructed) forms and meanings into account. Thereby, it is ensured that the semantic relations investigated are limited to words that are etymologically directly related, as intermediary steps may be lost (Blank 2001: 17-19) in completely synchronic data. For instance, in the example above, I would study the historical changes seen in Figure 1, yet not the purely synchronic relation between modern German Glück ‘good fortune, happiness’ and Unglück ‘bad fortune’, as the latter is not etymologically derived from the modern meanings of the former.
In the following, I first describe the data and the semantic and morphological categories used in this study and provide an introduction to how these categories fit into a usage-based model of the creation of lexical meaning (Chapter 2). Then I give an overview on the results of the study in regards to the various interactions between semantic, morphological, and other aspects of the data (Chapter 3). Chapter 4 discusses what conclusions can be drawn based on the results. The paper ends with a section on limitations and possible future studies (Chapter 5).

2. Materials and Methods

2.1 Materials

The data set used in this study consists of 480 German nouns. It is derived from the nominal entries in the German word list of the NorthEuraLex project (Dellert et al. 2020), based on the online version at http://northeuralex.org and an enhanced version by the Lexibank project (List et al. (forthcoming)). The concept-based presentation of the data enables evaluating the data by semantic fields and re-usage of the results in cross-linguistic studies. This is possible also beyond comparisons between NorthEuraLex data sets since the concepts are mapped to the Concepticon (List et al. 2021), which links them to a large range of other data sets. This list was chosen due to its relatively large size, its focus on basic vocabulary, and its overall reliability.

The word list was restricted to nouns because semantic relations have been studied particularly for this part of speech, so that for the categorizing used in this study I could build on previous work. A German word list was chosen because well-established etymological information is available for this language.

The data was enhanced using etymological information primarily from the etymological dictionaries by Kroonen (2013) and Pfeifer (1993), with additional consultation of Rix (2001), Wodtko et al. (2008), Köbler (2014), Lexer (2021) and the Frühneuhochdeutsches
The etymologies were in principle accepted as is. In line with Rix (2001: 6f.), root extensions, i.e. “synchron funktionslose[r] Morphe” (‘morphs without synchronic function’, ibid. 4f.), were not included in the Indo-European reconstructions in the data set. In general, reconstructions deemed uncertain in the etymological dictionary in which they were found were also not included. Thereby, the most uncertain semantic connections are avoided. For many words, a reconstruction to a not attested language stage was not necessary as the etymologies were restricted to the most recent word formation or borrowing event, i.e. to the moment that the word entered the German language or its ancestors. This ensures that the data remains limited to the 480 concepts and therefore comparable to other studies using this set of concepts.

As the focus of this study is the semantic side of the lexemes, the formal side, e.g. the orthography, has not been consistently harmonized between the sources that the forms were taken from.

2.2 Methods

This study uses a digital model of etymologies, building on the annotation framework developed in Schweikhard and List (2020) and Schweikhard and List (forthcoming), inspired by the standards proposed by the CLDF-initiative (Forkel et al. 2018). For this purpose, the etymologies were broken up into individual steps of the morphological, stratic and/or semantic processes presented below (see Figures 2 and 3), so that each such process could be analyzed separately. This approach enables the data to be evaluated semi-automatically, checked for consistency, and reused in other studies. Both the data and the scripts used for analyzing and consistency-checking are included in the supplementary material. Table 1 and 2 illustrate how the example from Figure 1 is represented in this method.

Table 1: Each formal or stratic process of each word’s history is represented in a spreadsheet, specifying the source and target forms and the semantic relation.

<table>
<thead>
<tr>
<th>ID</th>
<th>SOURCE</th>
<th>FORM SOURCE</th>
<th>CONCEPT SOURCE</th>
<th>PROCESS</th>
<th>SEMANTIC RELATION</th>
<th>ID</th>
<th>TARGET</th>
<th>FORM TARGET</th>
<th>CONCEPT TARGET</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>gelücke</td>
<td>good or bad</td>
<td>fortune</td>
<td>sound change</td>
<td>taxonomy</td>
<td>2</td>
<td>glück</td>
<td>good fortune</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>glück</td>
<td>good or bad</td>
<td>fortune</td>
<td>sound change</td>
<td>contiguity</td>
<td>3</td>
<td>Glück</td>
<td>happiness</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>gelücke</td>
<td>good or bad</td>
<td>fortune</td>
<td>derivation</td>
<td>taxonomy</td>
<td>4</td>
<td>ungelücke</td>
<td>bad fortune</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ungelücke</td>
<td>bad fortune</td>
<td>sound change</td>
<td>unchanged inheritance</td>
<td>5</td>
<td>Unglück</td>
<td>bad luck</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In order to categorize semantic relations, I use a two-dimensional grid proposed by Koch (2001), including both morphological and semantic relations.

Koch (ibid.: 1162) presents a scale of morphological similarity, beginning with full identity (i.e. polysemy or homophony), continuing with various forms of word formation patterns, from tone alternation to affixation, composition and serial verbs, and ending with lexicalized syntagms and idioms. These formal categories are strongly language-dependent and frequently even co-occur (e.g. when tone alternation and affixation are simultaneously involved in a derivation). Furthermore, the transparency of a pattern of e.g. stem alternation depends on how productive it is in a given language, so that not only pure objective similarity
but also similarity to patterns in a language’s lexicon would need to be taken into account to fully use this scale. Therefore, for the current study, I only distinguish between compounding on the one hand and other forms of word formation on the other hand. Compounding here refers to the combination of two otherwise free morphemes, whereas derivation includes affixation and internal derivation.

Table 2: Additional information for each word like the part of speech are provided in a separate spreadsheet. Identifier numbers link the words to the spreadsheet of processes.

<table>
<thead>
<tr>
<th>ID</th>
<th>LANGUAGE</th>
<th>FORM</th>
<th>PART OF SPEECH</th>
<th>CONCEPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Middle High German</td>
<td>gelücke</td>
<td>noun</td>
<td>good or bad fortune</td>
</tr>
<tr>
<td>2</td>
<td>Early Modern German</td>
<td>glück</td>
<td>noun</td>
<td>good fortune</td>
</tr>
<tr>
<td>3</td>
<td>German</td>
<td>Glück</td>
<td>noun</td>
<td>happiness</td>
</tr>
<tr>
<td>4</td>
<td>Middle High German</td>
<td>ungelücke</td>
<td>noun</td>
<td>bad fortune</td>
</tr>
<tr>
<td>5</td>
<td>German</td>
<td>Unglück</td>
<td>noun</td>
<td>bad fortune</td>
</tr>
</tbody>
</table>

Besides these word formation processes, also other processes can affect the formal side of a word. Here, I distinguish between analogical contamination, levelling, and inheritance from one language stage to the next, which often involves regular sound changes. Additionally, I also include the stratic dimension in the form of borrowing. For an overview on the processes included, see Figure 2.

Koch (ibid.: 1143) distinguishes between three semantic or associative relations in the lexicon: contiguity, similarity, and contrast. Similarity is split into metaphor and taxonomy, and taxonomy is further differentiated into hyponymy, hyperonymy, and co-taxonomy. This leads to six possible relations, plus semantic identity as a seventh possibility.

Metaphor describes that meaning 1 and meaning 2 have some characteristic in common. For instance, an ‘elbow’ has its shape in common with a ‘bow’. Hyponymy and hyperonymy
refer to the situation that meaning 1 is a subset of meaning 2, or vice versa. In the example of a meaning ‘bad fortune’ being derived from a meaning ‘good or bad fortune’, or in that meaning turning into only ‘good fortune’, the semantic relation is hyponymy as it refers to a narrower concept. In co-taxonomy, both meaning 1 and meaning 2 are subsets of the same meaning 3. Contiguity describes an (actual or perceived) local, temporal or logical relation between meaning 1 and meaning 2. It primarily includes the relation between a frame and its elements, or between elements of the same frame, for example part-whole relations (ibid.: 1145). One example of causal contiguity is the relation between ‘good fortune’ and ‘happiness’, the latter being caused by the former.

While Blank (2001: 17f.) states that similarity and contiguity can be combined in a single direct relation for some meaning pairs, I categorize each pair in the data only as one kind of relation. I go by the principle that taxonomic relations are the narrowest category (it can usually be unambiguously stated whether something is taxonomically related to another meaning) and metaphors are the broadest as some form of similarity can be found between most things. Therefore, contiguity relations are considered when taxonomy could be excluded, metaphor only when also contiguity could be excluded, and in one case, no semantic relation could be determined.

For further sub-categories of contiguity relations between nouns, I refer to the Bourque2 model in Pepper and Arnaud (2020: 121), which was developed to compare the semantic relations between the head and modifier of compounds. It distinguishes primarily between relations of containment and relations of causation. I use this model because it facilitates determining contiguity by providing clear descriptions of its sub-types. Not all of its sub-types turned out to be relevant for this study, though, and in the evaluation, I only distinguish between the primary sub-types of containment and causation. An overview on all semantic relation types distinguished in the study can be found in Figure 3.

Regarding semantic relations between other parts of speech and nouns, I make use of the categories in Bierwisch (2015) and Comrie and Thompson (2007). Semantic relations affect different parts of speech differently due to the different kinds of ontological types usually expressed by each part of speech (Bierwisch 2015: 1068). This study only includes nouns, which tend to refer to objects or substances (ibid.), but which may be derived either from nouns or from other parts of speech (primarily from verbs or adjectives, but also adpositions, adverbs or numerals, and even from phrases).

Figure 3: Overview on the categories of semantic relations with examples based on the concept ‘bow’.

For further sub-categories of contiguity relations between nouns, I refer to the Bourque2 model in Pepper and Arnaud (2020: 121), which was developed to compare the semantic relations between the head and modifier of compounds. It distinguishes primarily between relations of containment and relations of causation. I use this model because it facilitates determining contiguity by providing clear descriptions of its sub-types. Not all of its sub-types turned out to be relevant for this study, though, and in the evaluation, I only distinguish between the primary sub-types of containment and causation. An overview on all semantic relation types distinguished in the study can be found in Figure 3.

Regarding semantic relations between other parts of speech and nouns, I make use of the categories in Bierwisch (2015) and Comrie and Thompson (2007). Semantic relations affect different parts of speech differently due to the different kinds of ontological types usually expressed by each part of speech (Bierwisch 2015: 1068). This study only includes nouns, which tend to refer to objects or substances (ibid.), but which may be derived either from nouns or from other parts of speech (primarily from verbs or adjectives, but also adpositions, adverbs or numerals, and even from phrases).
Noun to noun derivations form the largest group of word formations in the data set. The categorization of semantic relations between source and target word of this derivation type does not differ from the one between the old and new meanings of a semantic change.

In verb to noun derivations, the respective semantic change is often either an event nominalization (Bierwisch 2015: 1073) or based on an argument (Comrie and Thompson 2007: 4-15). Argument-based nominalizations form contiguity relations because they refer to specific thematic roles of the action expressed by the verb, rather than to the action as a whole. “Verbs surrounded by their arguments” form a frame (Koch 2016: 46), changes affecting the argument roles are therefore based on contiguity (ibid.: 54). These verb nominalizations often refer to a prototypical instance of the respective thematic role of the action. I distinguish between the purely grammatical arguments of subject and object, and the thematic roles of location, time, instrument, causer, and result. For subject and object, I further distinguish between past and non-past (see also Comrie and Thompson 2007: 19). In the evaluation, these sub-categories were not considered due to the limited amount of data.

Event nominalizations on the other hand constitute an identity relation by way of merely expressing the same meaning via a different part of speech.

Change of the argument-independent core would also be a possible semantic difference (Koch 2016: 54) and could therefore include metaphorical or taxonomic relations. It is not found in my data set, likely because only changes from verb to noun, not from verb to verb are included.

Similarly to event nominalizations based on verbs, adjective to noun derivations can be state nominalizations (Bierwisch 2015: 1077), which thereby constitute an identity relation in the same fashion. Further sub-categorizations would be possible: The resulting noun can refer to the absolute state, the relative degree, or even a concrete object that is entirely defined by being in the state expressed by the adjective, rather than being something else that merely has that attribute usually or prototypically.

Frequently however, the meaning of adjective to noun derivations rather refers to an object which has the quality expressed by the adjective as a prototypical attribute, thereby forming a kind of contiguity relation. In some instances (only as the modifier of compounds in the data set studied here), it is a metaphorical sense of the adjective meaning which is an attribute of the resulting noun.

Compounds in the data set include noun + noun, adjective + noun, adposition + noun, adposition + verb, and verb + verb. Where adpositions were involved, only the relation between the other component and the whole compound was included in the overall evaluation. Due to their small amount, relations involving any of the three adverbs, nine adpositions and one phrase in the data set were not further categorized and the one numeral was counted as an adjective.

In some instances of reconstructed forms, the part of speech of the source form could not be determined with certainty. In all those cases, also the meaning was too uncertain so that those instances were not included in the overall evaluation, as were any other relations involving a too uncertain reconstructed meaning.

Besides semantic change based directly on the former meaning, also two semantic loans were found where the new meaning was derived from a word of another language.

2.3 Semantic aspects of language change

The interpretation of the results is based on pre-existing theories of the mechanisms of semantic change and of a usage-based explanation of language change. My primary references are Enfield (2015) and Koch (2016).
Enfield (2015: 3) states that words have a core meaning, which can be extended in conversations, using specific mechanisms. This means that innovation happens “at the discourse-level” (Koch 2016: 27), where it is guided by “inferences, implicatures and relevance” (ibid.: 28). Extensions are made not for the purpose of creating a new core meaning but for the communicative purposes within a given communication (Enfield 2015: 20). It is when the hearer assumes the extended meaning to be the core meaning that the word starts to acquire a new meaning. Meaning is created and spreads in the way it is understood by hearers in actual language usage (ibid.: 27) because the representations of meanings in the individual mental lexicons can only ever be hypotheses about the actual meaning (ibid.: 11) i. e. some intersubjective average which does not really exist (ibid.: 154).

Besides this hearer-driven spread of speaker-generated innovations, also hearer-generated innovations are possible. This is the case in ambiguous contexts or constructions, a so-called “bridging context” (ibid.: 26) where the hearer analyses what seems to be meant in the situation as the conventional meaning (Koch 2016: 30, Ullmann 1972: 195), without the speaker intending an extension of the meaning.

Koch suggests that contiguity-based semantic change often starts with the hearer (Koch 2016: 45). If a word is used frequently in a specific context, the hearer assumes that the respective frame, or rather a contiguity relation between the word and the frame, is inherent in the word (ibid.: 35).

Word formation almost always involves an innovation by the speaker (the only exception being univerbations, i. e. strings of morphemes which were produced as separate words but analyzed as a single word). Semantic change on the other hand can be caused either by meaning extension, i. e. an innovation by the speaker, or by reanalysis or misunderstanding in ambiguous contexts, i. e. an innovation by the hearer. Therefore, semantic relations between the source and target word of word formation processes show us what kinds of semantic relations play the biggest role in speaker-based semantic extension, whereas the semantic relations in semantic change include a mix of both speaker- and hearer-created innovations.

As a third central process, borrowing (according to Koch 2016: 56), is usually only hearer-based, thereby potentially showing what relations are involved in understanding rather than
producing. Figure 4 provides an overview on the proposed interaction between the primarily process types and the speaker or hearer.

3. Results

3.1 Overall frequency of processes

In the morphologically and semantically attested or reconstructable word history of the 480 nouns, 255 instances of word formation processes were found (36 compoundings, 219 derivations). This means, 53.1% of the sample are known to be directly derived from one or two specific other words with known meaning within the same language. Other morphological processes like analogical contamination and levelling were expectably rare, at 7 and 9 instances (1.5% and 1.9%), respectively. 76, or 15.8%, of the sample were borrowed from another language.

In the whole sample, 403 analyzable instances of semantic relations were encountered, including 2 instances of semantic loan, 13 cases involving adpositions, adverbs, or phrases, and 21 instances of an identity relation between source and target of a noun to noun word formation process, i. e. full homonymy.

50.2% of the words in the sample have only one semantic change in their reconstructable history, 12.9% have two semantic changes, and one word (0.2%) has three. For 36.7%, no semantic change is clearly reconstructable\(^1\).

Disregarding the loans and the relations involving adpositions, adverbs, or phrases, of the remaining 367 semantic relations, 258 were encountered as part of a word formation process, 94 as part of semantic change, 14 in borrowing events, and 1 as a side effect of analogical contamination. The following results only include semantic processes between the main parts of speech.

3.2 The influence of reconstruction

A caveat in the interpretation of the results is the fact that the data contain reconstructed meanings. Therefore, I compare semantic relations including reconstructed meanings against those which only include attested meanings. Reconstructedness has a drastic effect on the frequencies of semantic relations. Among relations between nouns, the relative amount of taxonomic relations is almost halved in reconstructed relations compared to unreconstructed ones, and also contiguity is slightly reduced, whereas metaphor is almost twice as frequent and identity around four times (Figure 5). I assume that this is an artifact of the process of reconstruction. It would imply that semantic reconstruction tends to overlook taxonomic and contiguity-based changes, instead opting for positing the same meanings, and that metaphor is overused as a semantic link in etymological reconstruction.

Adjectives show the same trend in regards to contiguity and identity but the results are less significant due to the smaller sample size. Here, metaphors are only found in attested data, which is likely caused by the increased frequency of compounds among attested data. Relations between verbs and nouns do not show any effect of reconstructedness due to the limited amount of relation types found there.

Also the frequency of the formal and stratific processes overall and in combination with semantic change differs depending on attestation status (Figure 6). Unsurprisingly, borrowing is found almost exclusively in the attested data, as it is very hard to reconstruct a borrowing from a not attested language stage. More interesting is the frequency of derivation, twice as

\(^1\) The amount of changes per word is higher by 1 than the amount of changes in the data set due to a pair of homonym words in the sample, which share one semantic change.
high in reconstructed data. This may have to do with the longer time span covered from Old High German to Indo-European as opposed to from modern German to Old High German, or it points to the reduced reliability of the reconstructions. Possibly, a certain part of the unattested derivations, particularly those with identical meaning between source and target form, may actually rather be instances of change of inflection class, or other forms of levelling or analogy, also explaining the high amount of identity relations.

Compounds are significantly more common among the attested part of the data set as well. This is again likely related to the process of reconstruction, in which compounds are rarely reconstructed as a whole, unless they lost their transparency to sound changes.

Figure 5: Frequency of semantic relation types depending on attestation status and parts of speech.

Figure 6: Frequency of morphological and stratic processes depending on attestation status.
Due to the potential shortcomings of the reconstructed data, I will base the evaluation primarily on the attested data.

3.3 Differences between parts of speech in word formation

As could already be glimpsed from Figure 5, part of speech has a strong influence on the semantic relations found. Figure 7 only includes word formation processes, as borrowing and semantic change only occur between the same parts of speech, i.e. between two nouns.

Contiguity is the most frequent kind of relation in all cases (except for reconstructed nouns), yet the complete absence of other relations besides identity (i.e. state or action nouns) and, for adjectives, metaphors shows that it is not comparable. The different parts of speech therefore need to be interpreted separately.

It can already be stated nonetheless that there are some commonalities across the whole data. In no category, relations of contrast were found, and also co-taxonomic relations turned out to be very rare.

It can already be stated nonetheless that there are some commonalities across the whole data. In no category, relations of contrast were found, and also co-taxonomic relations turned out to be very rare.

3.4 Semantic relations in different processes among nouns

Next, I look at the influence of different processes (different kinds of word formation, semantic change, borrowing, or analogical contamination) on the semantic relation type (see Figure 8). Only semantic change, borrowing and the different kinds of word formation include enough data points of semantic relations to compare different kinds of relations. In levelling, semantic change is completely absent, and for analogical contamination there is only one case, and only in reconstructed data. In order to compare between different process types, I only look at noun to noun relations, as relations involving other parts of speech normally only include word formation processes. I also exclude identity as a relation as it is an actual relation only in word formation, whereas in semantic change and borrowing, it is the default. Among nouns, there are 6 (9.5%) attested and 15 (35.7%) reconstructed instances of identity relations in word formation (primarily in derivation), and 61 instances (81.3%) of borrowing without semantic change. This means, as expected, borrowing usually does not cause semantic change, whereas word formation usually does.

In the attested data, taxonomy is the most frequent kind of semantic relation among nouns overall, yet not in all kinds of processes and only slightly behind contiguity. The sub-categories of hyponymy and containment are clearly favored over those of hyperonymy and causation, but also this depends on the type of process.

In word formation, containment-based contiguity is most common, found primarily in derivation and between the modifier and the whole compound in compounding. Nominal compound modifiers in particular exhibit only contiguity relations in the attested data.  

Figure 7: Frequency of semantic relations in word formation processes depending on part of speech. (This and the following figures contain only fully attested data).
The head of compounds on the other hand is in the majority of cases a hyponym of the whole compound, yet also various other relations are found here. Hyperonymy is completely absent from compound heads, which does not seem surprising. Metaphoric relations are not found in attested derivations or compound modifiers, which might be due to the small sample size.

Figure 8: Frequency of semantic relation types depending on the morphological or stratic process and in semantic change.
In semantic change, the contiguity relation is also frequent, yet not as frequent as in word formation, particularly if only attested relations are considered. Causation-based contiguity is slightly more common here than containment-based contiguity. In the attested part of the data, taxonomic relations, particularly hyperonymic ones, are most frequent. All kinds of semantic relations found in other process types are also found here, with the exception of co-taxonomy.

Semantic difference in borrowing is found almost only in attested instances. Even more so than semantic change, borrowing favors hyperonymic relations. Yet all major kinds of semantic relations are possible here. There were no examples for co-taxonomy and causation-based contiguity, possibly again due to the small sample size.

3.5 Semantic relations in word formation from adjectives and verbs

Word formation between different parts of speech clearly drives contiguity relations (Figure 9). Derivations also show a certain amount of identity relations, i.e. event nominalizations or state nominalizations. State nominalizations are quite frequent, accounting for 56.3% of derivations from adjectives. Metaphorical use of adjectives is only encountered in compounds, and here only in the modifier, which might be due to the fact that there are no instances of compounds with an adjective as the head.

Figure 9: Frequency of semantic relations in word formation from adjectives and verbs. There is no data for adjectival compound heads or verbal compound modifiers.

All instances of verbs being part of a compound involve either an adposition or, in one case, another verb, as the other constituent, and except from one example, they are all reconstructed. They furthermore could also be interpreted as not being derived directly from verbs but from a nominalization of the respective verb. This interpretation would slightly increase the amount of contiguity relations in the compounding of nouns, particularly in the unattested cases, which however wouldn’t change the overall results.
3.6 Influences of the semantic field

Basic vocabulary is not a strictly defined collection of concepts and a word list may or may not be balanced across different semantic fields. Such a balance would be difficult to ensure in any case as also semantic fields are to a certain degree a matter of arbitrary categorization and selection. Even for something as seemingly obvious as parts of the human body, many difficulties can be encountered. These include cross-cultural differences in the borders between body parts, and different degrees of salience and human affordance between for example ‘lungs’ or ‘hand’ on the one end, and ‘spleen’ on the other end.

<table>
<thead>
<tr>
<th>Semantic Field</th>
<th>Attested</th>
<th>Reconstructed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objects (N=329)</td>
<td>analogical contamination 0.9</td>
<td>analogical contamination 0.3</td>
</tr>
<tr>
<td></td>
<td>borrowing 11.2</td>
<td>borrowing 0.9</td>
</tr>
<tr>
<td></td>
<td>compounding 1.8</td>
<td>compounding 1.5</td>
</tr>
<tr>
<td></td>
<td>derivation 10.6</td>
<td>derivation 26.4</td>
</tr>
<tr>
<td></td>
<td>levelling 0</td>
<td>levelling 1.2</td>
</tr>
<tr>
<td>Animals (N=39)</td>
<td>analogical contamination 2.7</td>
<td>analogical contamination 2.6</td>
</tr>
<tr>
<td></td>
<td>borrowing 0</td>
<td>borrowing 0</td>
</tr>
<tr>
<td></td>
<td>compounding 2.7</td>
<td>compounding 0</td>
</tr>
<tr>
<td></td>
<td>derivation 7.7</td>
<td>derivation 25.6</td>
</tr>
<tr>
<td></td>
<td>levelling 0</td>
<td>levelling 0</td>
</tr>
<tr>
<td>Body Parts (N=48)</td>
<td>analogical contamination 0</td>
<td>analogical contamination 2.1</td>
</tr>
<tr>
<td></td>
<td>borrowing 6.3</td>
<td>borrowing 0</td>
</tr>
<tr>
<td></td>
<td>compounding 8.3</td>
<td>compounding 0</td>
</tr>
<tr>
<td></td>
<td>derivation 0</td>
<td>derivation 39.6</td>
</tr>
<tr>
<td></td>
<td>levelling 4.2</td>
<td>levelling 2.1</td>
</tr>
<tr>
<td>Persons (N=33)</td>
<td>analogical contamination 0</td>
<td>analogical contamination 0</td>
</tr>
<tr>
<td></td>
<td>borrowing 18.2</td>
<td>borrowing 0</td>
</tr>
<tr>
<td></td>
<td>compounding 12.1</td>
<td>compounding 0</td>
</tr>
<tr>
<td></td>
<td>derivation 3.0</td>
<td>derivation 24.2</td>
</tr>
<tr>
<td></td>
<td>levelling 0</td>
<td>levelling 3.0</td>
</tr>
<tr>
<td>Time (N=31)</td>
<td>analogical contamination 0</td>
<td>analogical contamination 0</td>
</tr>
<tr>
<td></td>
<td>borrowing 41.9</td>
<td>borrowing 0</td>
</tr>
<tr>
<td></td>
<td>compounding 22.6</td>
<td>compounding 0</td>
</tr>
<tr>
<td></td>
<td>derivation 3.2</td>
<td>derivation 6.5</td>
</tr>
<tr>
<td></td>
<td>levelling 0</td>
<td>levelling 0</td>
</tr>
</tbody>
</table>

*Figure 10: Frequency of morphological and stratic processes in different semantic fields.*
In order to check for the influence of the semantic field on the semantic relation, I sorted the 480 concepts into four semantic fields of animals (39 concepts), body parts (48), persons (33), time (31), and a general category of objects (329). To make sure that only semantic changes into the respective semantic field are included, the data set was restricted to only the word histories since the most recent semantic change that lead to the concept under study. This reduced the total amount of changes from 403 to 335.

Due to the limited amount of data, the following conclusions are only tentative. It nonetheless could be noted that the semantic fields behave clearly differently in certain aspects.

In regards to the processes they undergo (see Figure 10), particularly the frequency of borrowing varied strongly, ranging from 0% in the field animal and 6.3% for body parts, to above average 18.2% among person concepts, to a striking 41.9% in the semantic field time. Derivation on the other hand was below average for all four semantic fields examined (between 0% for body parts and 7.7% for animals) among attested data. Reconstructed data shows a slightly different picture here, pointing to the different diachronic stability between semantic fields, so that body part concepts are actually derived in more than average frequency in the reconstructed parts of their histories. Compounds were particularly frequent among person terms (12.1%) and especially the field time (22.6%) and rare among animals (2.6%).

The frequency of semantic change for each those process types also is influenced by the semantic field: Borrowing in no case changes the meaning of a word in the categories of time and body part, yet in 2/3rds of the borrowing events of the person category. Therefore, depending on the amount of concepts included for each semantic field, very different results will be found for the frequency of semantic change and other processes.

Concerning the kinds of semantic relations relevant for each semantic field, also some tendencies can be encountered, e. g. the absence of metaphor relations for coining animal or time terms. The very small amount of data for each category does not allow for further conclusions, though.

4. Discussion

Overall, contiguity was shown to be the most important semantic relation also for diachronic data. This confirms the results of the synchronic study by Koch and Marzo (2007). However, it also is clear that the various processes utilizing these semantic relations behave in very distinct ways.

Noun to noun word formations do not semantically differ in principle from semantic change in the kind of semantic relations involved. Only the frequency of the types of semantic relations differs. Therefore, they provide the best clues in regards to what drives semantic change versus innovation of semantic meaning in word formation and differences in meaning during borrowing.

Comrie and Thompson (2007: 72) state that in nominal derivation, abstract nouns, which refer to one quality of the source meaning, and augmentative, pejorative and diminutive nouns, i. e. those referring to a specific version of the source meaning, would be most common. The former constitutes a contiguity relation, the latter various kinds of hyponymy. Hyponymy and various forms of contiguity play a role in derivation, yet also hyperonymy and even homonymy was found in the data.

Compounds are said to make a contiguity relation more explicit according to Koch (2016: 34f.) He provides as an example hunting dog, in which the contiguity relation between dogs and the hunt is made explicit by the modifier hunting added to the simplex dog. My results confirm that this is the most important function of compound modifiers.
Additionally, the data show that compounds often stand in a metaphoric or hyponymic relation to their compound head, confirming Koch’s (2001: 1151, 1155) hypothesis that subordinate terms are often compounds with their hyperonym.

As word formation is based on speaker-innovations, these results imply that speakers use all forms of semantic relations in coining new words, yet align the word formation method with the semantic relation.

Semantic change in borrowing as a hearer-based process on the other hand is not less limited in regards to the possible semantic relations. However, it seems to be the case that hyperonymy is primarily a semantic relation utilized by the hearer. Speakers try to increase the precision of their statements via coining more specific words based on hyponymy. Hearers, who can only estimate the meaning, assume a larger scope of possible meanings – or a smaller scope in the cases of hyponymy. A similar process could explain the containment-based forms of contiguity, which typically are part-whole relations.

These results contradict the prediction by Koch (2016: 55f.) that semantic change in borrowing, or hearer-based change in general, is driven by hyponymy and contiguity. It needs to be tested on a larger data set, though, as semantic change is fairly rare in borrowing in all categories.

Semantic change independent of morphological or stratic processes is expected to show a combination of the other two process types as it is caused both by speaker- and hearer-innovations. This interpretation is consistent with the results, with semantic change overall resembling borrowing more closely by sharing with it the high amount of taxonomy in general and hyperonymy in particular. As the hearer plays a role in both hearer- and speaker-induced semantic change, a stronger role of semantic relations associated with the hearer is not unsurprising.

Causation-based contiguity is higher in semantic change than containment-based contiguity, unlike both in word formation and borrowing, which could be further investigated on larger data sets.

The absence of the semantic relations of co-taxonomy and contrast from semantic change is in line with expectations as both tend to lead to communication problems (ibid.: 37, 51). This is obvious with contrast as it would imply that a word can also mean its own opposite. Co-taxonomic polysemy would be a bit more tolerable but would also be a very unstable state. Considering the rarity or absence of these relations from word formation is nonetheless striking and, if not an artifact of the selected data, points to a general low importance of these semantic relations. Contrast is probably rather expressed by other means like syntax and co-taxonomy might simply not be very useful for communication. The low amount of co-taxonomy in the data might however also be caused by difficulties differentiating it from metaphor, which however would only cause a marginal increase in instances.

5. Limitations and future studies

A wide range of factors were determined which influence semantic relations. As this is a study on one particular language, the results will need to be confirmed by further studies to find worldwide trends. Due to the onomasiological approach of the study, with the concept serving as a tertium comparationis (Koch 2001: 1143), direct comparison of the results with those of future studies using the same concepts is possible.

Not included in the analysis was the frequency of semantic relations at different language stages, thereby determining whether this changes through history. However, often it is not possible to pinpoint at which time a change happened. This particularly applies to reconstructed language stages, but even for attested language stages, the etymological information
provided is often vague and a study of text corpora would be necessary, which is beyond the scope of this paper.

Similarly, also the influence of the structure of the semantic field as a whole at any given point in time could not be included. Each semantic change and coinage was studied in isolation.

The use of only etymologically valid relations reduced the amount of subjectivity in finding semantic connections in the lexicon. Yet there remains an element of subjectivity in categorizing them into types of relations. This could be overcome by conducting psycholinguistic experiments on how language speakers categorize the relations between the concepts in question. Due to the highly specific concepts involved even in this data set of basic vocabulary, semantic relations determined in that way may not be as useful for cross-linguistic studies since additional connections would need to be included for each additional language. Therefore, an easier solution would be to have several researchers annotate the semantic relations independent from each other and then check for inter-annotator agreement.

Supplementary Material

I provide supplementary material in the form of data and code, hosted at https://osf.io/m7vu9/?view_only=447f423416b04ba999ff3d7811bc004a.

References


Blank, Andreas. 2001. Words and concepts in time: Towards diachronic cognitive onomasiology. metaphorik.de 1. 6-25.


http://www.koeblingerhard.de/ahd/ahd.html


https://www.woerterbuchnetz.de/Lexer

https://concepticon.clld.org/


https://www.dwds.de/d/wb-etymwb


