Metalinguistic comments and signals: What can they tell us about the conventionalization of neologies?

Abstract

Many neologies receive a large amount of metalinguistic focus during their conventionalization. This includes explicit metalinguistic comments, as well as several ways of emphasizing a new word qua word in running texts, so-called metasignals (e.g., quotation marks). This article reports from a large quantitative study of 360 Swedish neologies. It investigates the nature and the amount of metafocus during conventionalization. More than 96% of the neologies received metafocus at least once, but the mean proportion of metafocused citations was low, just under 3.5%. Metafocusing is likely to be more intense in early phases and is likely to decline over time. No long-term effects of metafocusing on the conventionalization process itself were found in corpus data.

Keywords: conventionalization, neologies, metalinguistic comments, metasignals, metafocus, corpus data

1. Introduction¹

We seem to have a tendency to comment upon new words:

- (1) *Göran Tägtström värjer sig inför uttrycket “**stafettläkare**” som används på många håll i landet. - **Stafettläkare** låter lite nedsättande. Jag föredrar att kalla mig konsultläkare.*

Göran Tägtström opposes the expression “**relay doctor**” which is used in many places around the country. - **Relay doctor** sounds a bit disparaging. I prefer to call myself consulting doctor.

¹ I would like to thank Gunnar Eriksson for various kinds of computer-related assistance and an anonymous reviewer for invaluable comments that have improved the text.

- (2) *Nordkorea har av USA brännmärkts som en "skurkstat", en benämning som Ryssland har avvisat.*

North Korea has been stigmatized by the USA as a “**rogue state**”, a designation which Russia has rejected.

- (3) *“Curlingföräldrar” är ett smittsamt och smart ord.*

“*Curling parents*” is a contagious and smart word [the word denotes parents who coddle their children and serve them eagerly, referring to the sweeping element of the game and implying that the parents sweep away obstructions for their children, who get excessive help and assistance]

These examples are from Svanlund (2009), a comprehensive analysis of the conventionalization process for nine Swedish compounds. A striking feature for most of these compounds was the large amount of metalinguistic comments they received. In all three examples above, we meet evaluative attitudes towards the words themselves. In (3) it is provided directly by the author, in (2) we are told that Russia does not accept a certain designation, and in (1) the evaluation stems from a doctor who opposes being called *stafettläkare*. This word is an unofficial term which denotes doctors who only work temporarily at a hospital or care center. They are usually succeeded by other temporary doctors, and the doctors are metaphorically compared to runners in a relay race. Note that the cited doctor’s claim about an evaluative tone attached to the word is also reported in a direct quotation as a motivation for his own opposition.

We also meet comments about the usage and diffusion of these words. In (1) it is said that the word is used in many places around the country, and something similar is implied by characterizing *curlingföräldrar* as contagious in (3). In example (2) we learn who is using the word and about whom. The verb *brännmärka* (‘stigmatize’) signals a specific naming perspective, i.e., that the word is used as a label by a certain naming agent (in this case USA) who puts this label on a specific referent (in this case North Korea).

Metalinguistic comments might concern almost any linguistic property of the words, such as their pronunciation, spelling, meaning, usage patterns, etymology, history of usage, stylistic value, etc. They might also concern evaluations of any of these properties. In the study of Svanlund (2009), some aspects of these metalinguistic

perspectives even had a strong impact on the list of the top collocations for these words (see below).

A much larger quantitative study was undertaken to see whether similar observations hold for neologies in general. The aim was to investigate the nature and the amount of metalinguistic commenting and signaling (see below) during early conventionalization generally and what we can learn from analyzing these metalinguistic perspectives.

Some metalinguistic aspects of communication have received a fair amount of attention. One aspect concerns metalinguistic awareness of linguistic entities, which has mainly been studied from a language-learning perspective by for example Bower and Kawaguchi (2011) and Berry (2005). An interesting study is presented by Semino (2016), who investigates the kinds of metaphors that are characterized as *mixed metaphors* in discourse, and the evaluations attached to this label. In pragmatics, attention has been drawn to various metalinguistic phenomena, including *metalinguistic comparison* (e.g., *George is more dumb than crazy*) and *metalinguistic negation* (e.g., *Bill didn't paint the house, he slapped it all over with cheap whitewash*). Both examples compare the aptness of two categorizing expressions for labeling a specific experience (see e.g., Morzycki 2011). Other examples of metalinguistic negation concern more formal aspects, like pronunciation: *He didn't order '[eɪ]pricots'; he ordered '[æ]pricots'* (Morzycki 2011: 76). This comes fairly close to some metalinguistic comments in the data for the current investigation. However, they are overall more concerned with the general usage of a word, rather than specific instantiations in specific texts, as in the examples above. But there is no sharp line between commenting the use in a specific utterance and making remarks about repeated use in several related contexts.

There is a large body of research on neologies and word-formation productivity, and some scholars have tried to pin down factors that support conventionalization of new words (some examples are Bauer 1983, 2001, Baayen 2009, Benczes 2006, 2010). Metalinguistic perspectives on conventionalization has attracted much less attention but is covered by e.g., Fischer (1998), Metcalf (2002), Schmid (2008, 2016), Kerremans (2015), Svanlund (2006, 2009).

Hans-Jörg Schmid has taken great interest in the conventionalization and entrenchment of new words, also discussing metalinguistic perspectives. Schmid (2008) states that less transparent formations are more likely to reach a level of conscious

processing than entirely compositional ones. Speakers who become aware of producing a novel form can make decisions about potential decoding difficulties for hearers. In that case, they tend to choose from a range of strategies to mark the newness of the word. These include metalinguistic markers like *what you may call* and *so-called*, use of definitions and explanations, inverted commas etc. The existence of these strategies indicates that writers are sometimes aware of the newness of nonce-formations (Schmid 2008: 16–17).

While reflexes of metalinguistic awareness (on behalf of speakers) is repeatedly evidenced in the study to be presented below, the precise nature of this awareness is not always self-evident, especially not when it is only indicated through markers like quotation marks (see below). What metalinguistic aspects are the speakers actually aware of? And what effects does this higher degree of awareness have on linguistic processing and the conventionalization process?

2. Metacomments and metasignals

A comment or statement is here treated as *metalinguistic* if it pertains to linguistic properties of lexical items, i.e., focusing aspects of the code rather than (exclusively) predicating something about the extralinguistic reality. As exemplified above, such properties include for instance pronunciation, spelling, meaning, usage patterns, etymology, and history of usage. Metalinguistic comments can also relate to evaluations of any of these properties or of the term as a whole and its usefulness. From now on the shorter variant term *metacomment* will be used, and *meta-* in compound words is generally short for *metalinguistic* in this text.

Metacomments are often used together with markers which can be labeled *metasignals*. These signals draw the reader's attention to the word, often in order to comment upon the word itself. Svanlund (2009) distinguished four main types of metasignals, but only three of them are analyzed here, since they potentially provide the best precision and recall in corpus searches. They are all used to emphasize the word qua word in some respect. *Hedges* consist of two subtypes that are partly used in similar ways. One is constituted simply by quotation marks around the word, as in example (4) and in (1)–(3) above (sometimes referred to as “scare quotes”, see e.g., Predelli 2003);

the other consists of the Swedish phrase *så kallad* ('so-called', 'as it is called'), as in examples (5) and (6).

- (4) *Allt fler distriktsläkare säger upp sig. De blir "stafettläkare" och hyr ut sig själva för betydligt högre lön.*

More and more district doctors are resigning. They become "relay doctors" who take temporary posts for a considerably higher salary.

- (5) *Så kallade **curlingföräldrar** är föräldrar som sopar banan framför sina barn.*

So-called **curling parents** are parents who sweep the sheet [i.e., the playing surface] for their children.

- (6) *Tvåans "Reportrarna" handlade om de så kallade "**bokstavsbarne**n"--de som fått diagnoser som DAMP, MBD med mera, alltså störningar i hjärnan.*

TV2's "The Reporters" dealt with the so-called "**letter children**"--those with diagnoses like DAMP, MBD and so forth, that is disorders in the brain.

These two metasignals can also be used together, as indicated by example (6). They are often characterized as hedges, but hedging is only one of several functions that these markers fulfill. Both types can indicate that the word is being introduced to the reader, or that it is odd or contentious, or that the writer wants to dissociate himself or herself from the actual wording. One difference between them is that 'so-called' always appeals to some already established usage, some kind of convention, while quotation marks can be used also for completely nonce formations, which are sometimes the writer's own innovative coinage. Quotation marks also have several additional functions, where one of them, naturally, is to mark direct quotations. They can also signal for instance that a word is an unintegrated loanword, or that the chosen word is an approximation, or that it is currently being used in a deviant way, for example ironically, or stylistically inappropriate.

In example (7) quotation marks are used to indicate pronunciation, while hyphens between every letter of the word indicate its spelling:

- (7) *Dagens matsida börjar med en språklektion. Det stavas **b-r-u-s-c-h-e-t-t-a** och uttalas “brusketta”. Inget annat! H:et ger ett hårt k-ljud – utan återvändo. Så är det bara. Lägg det hemska gärna på minnet.*

Today’s food page starts with a language lesson. It is spelled **b-r-u-s-c-h-e-t-t-a** and it is pronounced “brusketta”. Nothing else! The *h* yields a hard *k*-sound – there is no turning back. That’s just the way it is. Please keep that in mind.

Word-pointers are expressions like “the word X”, “the term X” or “the concept X”. They point out a certain word in order for the writer or speaker to explicitly say something about it:

- (8) *Plötsligt har begreppet “**bokstavs**barn” fått en nedsättande klang.*

Suddenly the concept of “**letter child**” has assumed a disparaging tone.

- (9) *Många svenskar har i år införlivat ordet “**curling**förälder” med sitt ordförråd.*

Many Swedes have incorporated the word “**curling parent**” in their vocabulary this year.

The word itself may be controversial. Often the phenomena it denotes--the referents--are also hotly debated. Metacomments are often about the tone, attitude and values that are conceived to be attached to the new word, as in example (8). Sometimes the writer wants to convey his or her own evaluation of the expression, as in example (3) above. Sometimes the comments are more about the usage, the conventionality status, the diffusion or sometimes the success of the new word, as in (9).

It is rather unnatural--and therefore rare--to use word-pointers without adding an explicit comment, while hedges can be--and in fact often are--used in isolation, in which case they function as implicit metalinguistic statements by themselves. The overall function is something like “Watch out! There is something tricky about this word”. But the nature of this trickiness is frequently left implicit. Quite often it is related to conventionality status; the word is not assumed to be familiar enough for the

readers as yet. Quite often it also remains unclear what aspects, precisely, the writer wants to draw attention to. All this suggests that it can be useful to distinguish between *metasignals*, such as quotation marks, and explicit *metacomments*, but both are of interest when analyzing the conventionalization of neologies. They are here subsumed under the term *metafocusings*, which is the theme of this article.

Metasignals are sometimes used very frequently. In Svanlund's (2009) study of Swedish compounds in large corpora, some metasignals actually gained considerable collocational strength, and some collocating metasignals even belonged to the top collocates for the compounds under investigation--primarily the hedge *så kallad* ('so-called') and the word-pointer *begreppet* ('the concept').

The types of metasignals used in the present study are easily searchable with a high level of precision and recall. This is a great advantage when working with large corpora (see below). When the number of citations exceeds 500,000, it is virtually impossible to manually check each one for actual metacomments. Searching for metasignals offers a shortcut to finding metalinguistic statements, explicit or implicit. However, it turns out that many metacomments are never metasignaled at all. In Svanlund (2009), all citations were also manually tagged for explanations of the words' meanings, a specific kind of metacomment that is quite common for neologies (contrary to what would be expected if these were fully compositional, which is sometimes assumed to be the normal case). On average only one-third of the explained citations were also metasignaled (with a very large variance), while about one-fifth of the metasignaled citations contained explanations. This means that a large portion of this specific type of metacomment would be missed if we are forced to rely on only checking for metasignals. Hence, the present study attempts to supplement metasignals by at least retrieving a portion of the actual explanations through automatic search algorithms, irrespective of any metasignals. The search is based on certain frequent collocations that on testing were found to be highly associated with explanations, for instance the phrase *det vill säga* ('that is to say'), often abbreviated *dvs.* ('i.e.'). Thus, when we are quantifying metafocusings in this study, we are in effect counting citations with metasignals plus citations with explanations (in a broad sense), but no other types of explicit metacomments. The term explanation mostly covers the actual function, but the category is in fact a bit broader, pertaining to various kinds of semantic explications and comments.

There is no reason to posit a sharp dichotomous distinction between metalinguistic and objectlinguistic uses of a word, contrary to what is often assumed. One motivation for complicating this dichotomy comes from expressions of (negative) attitude. In many cases it is difficult to determine whether the attitude is directed toward the referent or the expression or both. There seems to be a tendency to direct one's anger against both the term and its reference, sometimes even without realizing this. One example concerns the term *trängselskatt* ('congestion taxes', 'congestion charges'), which denotes a fee that was introduced in some cities in Sweden in order to reduce traffic in city centers. Many citations bear witness to a critical attitude--both toward the term, which some find misleading, and toward the measures taken, which some find unnecessary or harmful. Sometimes there is also a shift in the same context:

- (10) *Flera av de närvarande vände sig mot uttrycket **trängselskatt** och menade att det var en skrivbordsprodukt för att finansiera Västlänken.*

Several of those present turned against the phrase **congestion taxes**, saying that it was a product from the drawing board, only designed for financing the Västlänken traffic route.

A linguistic expression can hardly finance a traffic route, but analogous contextual shifts are not rare. Correspondingly, when reference is made to how familiar a word or concept is, this frequently seems to involve both the term and the phenomenon it denotes, but very often it is simply unclear, and contextual shifts are abundant. Furthermore, explanations and notes about semantics are common types of metalinguistic comment. But it is sometimes difficult to tell if statements pertaining especially to people primarily concern the usage of words or the general whereabouts of their referents. For instance, are the examples (5) and (6) above primarily about actions typical of certain human categories, or are they explaining what certain words generally denote? Or perhaps both? In this text we sometimes present proportions of metalinguistic uses expressed in percentages. This is not meant to imply that these uses are *purely* metalinguistic and lacking any objectlinguistic function. It simply states that the given figure reflects the amount of uses interpreted to involve some kind of metalinguistic perspective, whether purely so or not.

3. Data

In this study, neologies are considered to be words that relatively recently have been introduced into the speech community. They have entered the process of becoming conventionalized to some degree, while many of them are still not familiar to large groups of speakers. In all 360 neologies from the major open classes were selected for analysis; i.e. nouns, verbs and adjectives. Most of them were picked from books and lists of new words published annually by the Language Council of Sweden² (about 50% of the words); some were taken from lists of updates to dictionaries, like SAOL (the Swedish Academy Glossary) (32%); and a few were added by myself, partly with some computational search assistance from Markus Forsberg at the Swedish Language Bank, see below (10%). Some of the items appeared in more than one of these sources (8%).

The selection deliberately reflects a large variety of origins, registers and word-formation types, as well as various levels of estimated transparency. Only new word forms were considered--new lexical items with earlier homonyms or earlier polysemously related senses were discarded.

Two important criteria for the selection concern age and frequency. Since one ambition was to study long-term effects of metafocusing, the selection had to include words of varying age. In fact a fair amount of the words has now reached considerable age, and some of them would hardly be described as neologies today. They are included in the selection since they were once explicitly acknowledged as neologies some twenty years ago. Since all their citations are included, from the very earliest until very recent ones, this makes it possible to track their evolution through time in a diachronic perspective. A time threshold was set for the inclusion of older neologies: the first attested citation should not be earlier than 1993. However, on closer examination it turned out that four of the selected words in fact were represented by a small number of earlier citations, with 1987 as the very earliest citation. In spite of this, they were kept in the final selection.

Varying age also suggests potentially varying degrees of conventionalization for the selected words, which was also the intention. Another factor often mentioned as

² One of the council's duties is to monitor the development of spoken and written Swedish and especially to register new words.

determining or accompanying conventionalization is frequency. The selected words purposely differ substantially in frequency. But in order to exclude pure nonce-formations, the selection was guided by estimations of the total number of available citations in the corpus resource Korp (see below), with a minimum threshold frequency of at least 100 tokens. This means that no hapaxes were included. All of the selected items are on their way of becoming conventionalized to some degree. However, during the analyses three words turned out to actually have less than 100 valid citations, with 61 as the lowest number. These words were also retained in the final selection.

All the available citations for the selected words were collected from the Korp interface at the Swedish Language Bank, a large collection of corpus resources managed by the University of Gothenburg.³ At this time (early 2014) Korp contained 144 subcorpora with a total of 1.78 billion tokens (i.e., words and punctuation marks). The corpora were classified into corpus types, mainly through the classification system used by the Swedish Language Bank itself. The largest types were newspapers, blogs and tweets, but for instance literature, academic texts and medical texts are also included. The total number of collected citations for the entire set of selected neologies amounts to 504,352. All examples given in this text are authentic citations from the corpus. For a few selected words, Korp turned out to contain some early citations for homonyms--mostly proper names--of the neologies. These early citations were discarded from the dataset and are not included in the figures presented here. No other “corrections” have been made to the included citations or the available taggings. It would have been too time-consuming to make adjustments to the entire dataset. This means that some apparently incorrect part-of-speech taggings are also preserved.

The most important research questions were formulated as follows:

- To what extent are neologies metafocused (i.e., explicitly called attention to through metasignals and metacomments)?
- What kinds of neologies are most metafocused?
- Can metafocusing in itself affect the conventionalization process? The hypothesis tested (which is more thoroughly presented in section 6) claims that metafocusing should speed up the conventionalization process.
- How can we use metafocusing in the analysis of neologies?

³ The Swedish Language Bank is available at <https://spraakbanken.gu.se/eng>.

Perhaps some notes on terminology and the trickiness of conventionality are in order here. *Neologies* and *new words* are used here interchangeably as synonyms. The noun *conventionalization* generally describes the process where a new linguistic unit becomes conventional. It is often used with the presupposition of an endpoint. When this is reached, the process is completed and the unit is simply considered *conventional*. (It is also often taken for granted that this will eventually happen.) *Conventionality*, however, is clearly a matter of degree--in many respects--and the “endpoint” is therefore not a question of all or nothing. Whether a linguistic unit is to be considered conventional or not is also closely related to which specific speech communities, registers, genres or other modes of communication we are talking about.

Conventionality status changes and sometimes it decreases. While it makes perfect sense to state that a unit, after some time, has become *less conventional* than before, it sounds more odd to say that it is now *less conventionalized*, since conventionalization usually refers to early stages--the onset--of a unit’s career, up to the “endpoint”. Nevertheless, *(less) conventionalized* is sometimes used also about decrease of conventionality status, regardless of when this occurs. It would perhaps generally be more preferable to talk about the *lexical strength* of lexical units in the sense of Bybee (1985, 2010), rather than the unit’s degree of conventionality or conventionalization; the advantage being that *lexical strength* carries no association to specific stages. In this text all the words under study will be treated as being involved in a *conventionalization* process throughout the studied time span, which is sometimes more than 20 years. This is not to be taken as implicitly stating that all of them actually will become fully conventional. Nor does it imply that they have *not* yet reached conventionality status by the end of the studied period. It could be argued that some of them were actually fully conventionalized already after, say, five years. But it would not be feasible to make continuous conventionality judgments for all the words over various time sequences. It is simply more convenient to consistently speak about their *conventionalization* process, disregarding assessments of how far they have come. But the reader should keep in mind the lack of presuppositions that this terminological choice is meant to carry in this text.

4. Patterns of metafocusing

A total of 9,430 metafocused citations were found for the 360 words. It is essential, however, to bear in mind that this figure does not reflect the entire number of metafocusings in the corpus, since only a single type of explicit metacomment was sought after (i.e., explanations), only a portion of this type could be retrieved with the search algorithms, and only three types of metasignals were used, leaving out for instance italicization.

The overall number of metafocused citations is presented in Table 1. Over 85% of the metafocused citations were metasignaled, 16% contained explanations, and less than 2% overlapped with both metasignals and explanations. Only 9% of the explanations were metasignaled (not presented in the table). If we only were to look for metasignals, we would certainly miss a substantial part of the explanations, one main type of metacomment. It is also worth noticing that about 4% of the metafocused citations contained more than one metasignal. However, the figures presented here refers to the number of metasignaled citations, not the number of metasignals per se.

Table 1. Citations and metafocused citations of 360 Swedish neologies in the Korp database

Number of	Adjectives	Nouns	Verbs	Total	Percentage of metafocusings
Neologies	25	293	42	360	
Citations	26819	396050	81483	504352	
Metasignals (ms)	612	6644	812	8068	85.6%
Explanations (exp)	128	1,291	78	1497	15.9%
Overlap ms & exp	7	117	11	135	1.4%
Metafocusings, total	733	7818	879	9430	

We can establish that neologies are indeed associated with metafocusing to a large extent. Over 96% of the new words were metafocused at least once (which is not indicated in Table 1, however). Metafocusing thus seems to be a normal phenomenon during lexical conventionalization. But for most of the words it is not an apparently dominant trait. The distribution of metafocusings is presented in Figure 1. The large majority of words have low values. The mean proportion of metafocused citations is only just under 3.5%, with a very large variance (the standard deviation is 5.4). The

mean is also affected by a few very intensely metafocused words (the small circles representing outliers in the figure), and the median is only 1.5%.

[Insert file SVA1 here, same as Figure 1 below]

Figure 1. Distribution of metafocusings

Topping the list of metafocused words is the massively metasignaled *otrohetskroll* ('unfaithfulness control'). It was spread in connection with a court case concerning a husband who used force to examine his wife's genitals in order to establish whether she had been unfaithful. Almost 47% of the citations were metasignaled--virtually always with quotation marks--which by far makes this the most metafocused word. None of the citations contained explanations, however, which indicates that the word was probably considered fairly transparent in context. Quotation marks here signal looseness and lack of official status for the term. The second most metafocused word is *skurkstat* ('rogue state') with 28% metafocused citations, containing both metasignals and explanations.

There is no simple overall correlation between metafocusing and standard conventionalization indicators such as frequency, age, or diffusion (measured here as the number of subcorpora or corpus types where the words are attested). Nor is there any straightforward association with factors indicating morphological complexity. But when we compare the most metafocused words with the rest, we find significant

differences. The most focused words are--on average--younger, less frequent and less diffused, which can be seen from Table 2.

Table 2. Common conventionality indicators, means for all 360 words divided into three groups

Variables	High-focus words (10%)	Middle-focus words (80%)	Low-focus words (10%)
	Mean	Mean	Mean
Metafocusings (%)	16.7	2.2	0.1
Metasignals (%)	15.0	1.8	0.1
Explanations (%)	4.7	0.7	0.0
Age (years)	11.1	15.2	16.2
Frequency (N)	344	1460	1984
Corpora (N)	24.9	39.0	47.0
Corpus types (N)	6.2	7.2	7.7

Table 2 shows average values for the whole set of neologies divided into three groups depending on their metafocusing values. The highest metafocus decile is represented by the left column. These are the 10% of the neologies that have the highest relative metafocusing values, i.e., the highest percentage of metafocused citations. To the far right, we have the lowest decile--the 10% that have the lowest relative values. In between those two is a column that represents the large bulk of the words, the 80% that lie between the two extreme groups. The table shows the average values for age, frequency, number of corpora and number of corpus types represented, along with the average metafocusing values. Thus, the high-focus group is on average represented in 6.2 corpus types, while the low-focus group is represented in 7.7 corpus types, which is indicated by the last row. The metafocusing values are separated into metasignals, explanations and the total metafocusing values.

All the variables in the table show a consistent pattern from the high-focus group through the middle-focus group to the low-focus group. The pattern is $1 < 2 < 3$, where the high-focus words have the lowest values and the low-focus words have the highest values (except for the metafocusing values in the first three rows where the order is reversed). All differences between the high-focus group and the rest of the words are statistically significant ($p < 0.001$ according to the Wilcoxon test with

continuity correction, which is the test generally used in this text unless otherwise stated, since virtually none of the variables are normally distributed). These results corroborate the conclusion that metafocusing is partly connected to novelty and lack of conventionalization status. Metafocusing is more used with less conventionalized lexical items. The highest metafocused decile contains the following 36 words:

- (11) *24-timmarsmyndighet, bloggövning, bokstavsbarn, budgetstup, bötning, curlingförälder, enveckasförsvär, EU-moped, fjärlseffekt, flärp, flexicurity, Foppatoffel, friåkare, grooming, jobbpakt, Juholtare, köttklister, mördarbakterie, multikulti, ogooglebar, otrohetskontroll, phishing, plusjobb, prekariat, risksupportrar, sjuknärvaro, skattekil, skimning, skurkstat, stjärnfamilj, storstadssatsning, svenskfientlig, svenskfientlighet, terja, triage, vobba.*

From this list of the 36 most intensely metafocused words, we can make some further observations that are also substantiated by analyses of the whole set of 360 neologies. Many of the most metafocused words are strikingly non-transparent. While transparency is a subjective experience on behalf of speakers--not an objective feature of words--it is fairly uncontroversial and intersubjectively safe to state that the high-focused group contains a large proportion of words that most people would consider semantically opaque. These words are not only morphologically complex (though in most respects the highest decile does not differ significantly from the rest of the words in complexity), they also rely on intricate metaphorical or metonymical mappings that have to be sorted out by the listener or reader. Figurativeness as such does not necessarily imply greater opacity, but complex mappings which are not based on well-established figurative patterns may lead to such effects. That is precisely what we see in some of these words. One example is *bokstavsbarn*, ‘letter children’, which refers to children with certain psychosocial problems. These children have received an intricate neuropsychiatric diagnosis which is usually referred to by an acronym of three or four letters, like *adhd*. Thus, the modifier appeals metonymically to a linguistic property of the diagnosis that characterizes these children. Some of the mappings also have a somewhat concealed humorous twist. By metasignaling these words, the author

acknowledges that the readers are not presumed to be familiar with the word or its meaning and are not presumed to be able to work this out by themselves.

The most metafocused words are also marginally longer, containing slightly more letters on average ($p < 0.05$). Other measures pertaining to complexity--like the number of syllables or morphemes--do not yield significant differences. It would indeed make sense if morphological complexity increased the need for explanations and perhaps also metasignals, but there is no straightforward support for this hypothesis in this study.

There are very few simplex words in the set (except for the large group of borrowings, see below), which makes them difficult to assess as a group.

Derivations are explained more seldom than neologies in general ($p < 0.05$). This includes Swedish derivations based on English borrowings, not included in the loanword category discussed below. Most of the derivational bases are already established lexical items. Since derivation is a more regular word-formation process than compounding, the need for explanations would generally be predicted to be small, which is also what we found.

The largest morphological word type in the set is compounds. With a total number of 203 compounds, this category represents more than half of the neologies studied here. Compounds lie very close to the average for both explanations and metasignals. But compounds are a heterogeneous group with a large variance. Morphologically complex words offer the listener or reader more semantically relevant material as input to the interpretation process. This could sometimes make it easier to generate a plausible interpretation, and new compounds could be experienced as quite transparent in context, which decreases the need for explanations. On the other hand, quite a few compounds are highly ambiguous, and some are based on advanced metaphorical and metonymical mappings as we saw earlier, which increases the need for metafocusing and especially explanations.

The 73 loanwords in our set have significantly more explanations than other words ($p < 0.001$), which may not be surprising. Many of them basically represent new morphemes to the Swedish language and should by definition be more opaque and difficult to interpret, and therefore in need of more explaining. But they are not metafocused more often, which means that they are not metasignaled to any extent. There are, however, major differences also within this large category. The English borrowings have significantly more metasignals than loanwords from other languages.

Especially foreign food terms have a lower degree of both metasignals and explanations. These are predominantly of non-English origin, such as *bresaola*, *ciabatta*, *gremolata*, *shiitake*, *haloumi* and *brioche*. There also seems to be a genre effect for food terms. In designated food-related contexts, it seems to be taken for granted that the reader expects to find some more or less exotic and more or less unfamiliar terms.

5. Decline over time

The citations for the 360 neologies represent a rather large span of time. The age of the neologies when the data collection was concluded in 2014 varied considerably--from 1 year to 27 years. Age was measured as the time distance in years from the date of the first citation in the corpus up to 2014. A large group of 88 words had by then been used for at least 20 years. This means that they would hardly still be described as neologies. It also means that we can follow their conventionalization patterns over 20 years of use. The time span makes it possible to answer questions about the evolution of metafocusing patterns over time. What we find is a clear general tendency of decline in metafocusing over time, which is indicated by Figure 2, covering the total set of 360 neologies.

[Insert file name SVA2 here, same as Figure 2 below]

Figure 2. Decline in relative metafocusing over time

In Figure 2 the time axis is presented as relative years, i.e., the number of years since the first citation in the corpus. We can see that the decline is not in form of a smooth line, but still evident. The means in early years are easily affected by extreme values of 50% or more, since the total number of citations is fairly low in early phases for some words, and it then only takes a couple of metafocusings to reach 50%. Since not all words are 20 years old, the total number of words reflected in the figure decreases for each year. If we only look at the 88 words that were at least 20 years old, we also find a similar pattern of decline in metafocusing (not presented in any figure here). One measure of this is that the mean proportion of metafocusing for those words decreases from 5.2% for the first 5 years to 2.2% during the last 5 years--a difference which is statistically significant ($p < 0.01$).

These figures all confirm that metafocusing is related to novelty to a large extent. When these neologies become more conventionalized, they are less likely to get metafocused. But a fair amount of them continues to receive metafocus for a long time, since metafocusing is by no means exclusively connected to novelty and conventionalization--e.g., technical jargon, obsolete concepts and controversial terms (relating to, e.g., race, gender or religion) may also motivate the use of metasignals and

metacomments. The status of being controversial may persist for some words, which can motivate the use of metafocusing for a long time.

This study has been unable to confirm Fischer's (1998) observation of changes over time in the strategies for explaining new words. The types of explanations used for the most commonly explained words in our data set are very much the same throughout the entire period.

6. How does metafocusing affect conventionalization?

One important question concerns what effects metalinguistic focusing may have on the conventionalization process itself. Two opposing views are discernable. Metcalf (2002) claims that metalinguistic comments restrain conventionalization.

The more general principle is that a word shouldn't call attention to itself, by humour or by any other means--its looks or origin, for example. If it does attract notice, chances are it will be rejected. If, however, it is unobtrusive and sounds familiar even if it is not, those who hear or see it are more likely to think it's something they have missed rather than something missing from the language. (Metcalf 2002: 144)

Metcalf's advice for those who want to plant a new word into the English language is "1. Camouflage it. 2. Smuggle it in. 3. Talk it up." (Metcalf 2002: 185). His views do not seem to be corroborated by systematic research, but rather stems from personal experience, for example from working on choosing the word of the year for the American Dialect Society.

An opposing view is stated by Schmid (2016), who claims that metalinguistic discussion promotes conventionalization. Words coined and used by people with prestige have better chances of diffusion, resulting in a higher probability that these words are not only reused but also become the topic of metalinguistic discussions. This in turn will contribute to their diffusion (Schmid 2016: 75).

Empirical investigations of the influence of metalinguistic comments are rare, however. A noticeable exception is Kerremans (2015), who tried to measure the influence of metacomments empirically in her study on 44 neologies. She found a marginal, but statistically significant, fueling effect of metalinguistic usage. A generalized linear binomial regression model, with overall frequency as the single dependent variable, resulted in 1.01 as the effect estimate factor, $p = 0.032851$. (Values above 1 indicate a positive effect, values between 0 and 1 indicate negative impact.)

In the present study, the hypothesis tested is in accordance with Schmid's reasoning and with Kerremans's findings: A large amount of metafocus--especially during early phases--should speed up the conventionalization process. This is built on the consideration that metafocusing makes words more noticeable, makes them stand out in texts, and may therefore cause longer processing times--and perhaps more intense processing--during reading. Therefore these words probably need less frequent exposure to become familiar, which in turn should speed up frequency growth and diffusion.

To test this, a similar analysis to Kerremans's was undertaken, with the time span of metafocusing and frequency development added as a dimension. The impact of metafocusing during early stages was modeled against frequency during later stages as the single dependent variable. The relationship was investigated through a series of generalized Poisson regression models covering the 88 neologies which had existed for at least 20 years by the end of the research period. Various divisions of time periods were tested, with for example the amount of metafocusing during the first five years compared to the total frequency during the last five years etc.

The models generally generate a highly significant ($p < 0.001$) but extremely weak negative effect of metafocusing. The effect estimate factor is usually just under 1, which indicates a negative effect. This would mildly contradict Kerremans's result, where she found an effect size of 1.01, a slightly positive effect. The effects in the present study would be expected from the results reported above, where highly metafocused words on average were found to be less frequent than the rest. However, the regression models are all heavily overdispersed, which indicates a large variance and a poor fit between observed and predicted values, which in turn implies that important factors are missing from a valid prediction of frequency developments. In other words, the relative number of metafocusings seems not to be a good predictor of success or failure in the conventionalization process.

The regression modeling was repeated for two other establishment indicators as response variables, namely the two dispersion measures: number of corpora and number of corpus types represented by the citations. The effects remain extremely weak and still slightly negative. These models generally suffer less from overdispersion, but on the other hand most of them are not statistically significant as a whole and neither is the effect of the metafocusing variable.

In sum, the hypothesis that metafocusing should promote conventionalization could not be confirmed by the data analysis. This leads to the cautious conclusion that the amount of metafocusing actually does not seem to affect the degree of conventionalization for neologies at all, at least not very much, and at least not in so far as it is quantitatively reflected in corpus data (which is often taken as self-evident that it should). The result from Kerremans (2105) could be interpreted in a similar way; the joint conclusion from these two studies is that the effect of metafocusing is marginal. This does not disqualify metafocusing from potentially having an impact on the qualitative nature of associations that we attach to these new words, for instance for highly controversial words denoting highly controversial phenomena, or for words which often have to be explained. It might also still be the case that metafocusing facilitates some aspects of cognitive processing, for instance the speed of word recognition. But that will be the target of another study. The precise effects on linguistic processing have to be assessed through psycholinguistic methods, and such an elaboration is under planning.

7. Conclusions

Metalinguistic comments and signals seem to be an integral part of the conventionalization process for most new words, but to varying degrees. This means that metafocusing can also be utilized to study aspects of this process. In one kind of application, metasignals are used as a way of computationally searching for potential neologies, exemplified by Kerremans's utilization of NeoCrawler, a tool which conducts searches for neologies on the web (Kerremans 2015).

Especially word-pointers can also be used to search for explicit comments about words undergoing conventionalization, which for some neologies would be an essential

trait. Some types of comments also provide valuable insights into the conventionalization process itself, or at least into how it is perceived by various users, and what semantic traits users consciously acknowledge. Word-pointers can for instance be used to trace (perhaps opposing) attitudes toward controversial words.

Since the amount of metafocusing generally seems to decline over time, this could be used as an indication of conventionality status along with more traditional measures of frequency and dispersion.

Finally, we should not forget that there are a host of reasons to comment upon words or to metasignal them. We should not expect all metafocused words to behave in a uniform way. Many neologies are not only new but also controversial and seemingly non-transparent, and perhaps they are also spelled or pronounced in a peculiar way, as epitomized by the troublesome *bruschetta* in example (7). Other words display just one of those traits--or they may have an interesting origin that induces comments and signals. Novelty changes over time by definition, while for instance controversy may very well persist.

References

- Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, 900–919. Berlin: Mouton De Gruyter.
- Bauer, Laurie. 1983. *English word-formation*. Cambridge: Cambridge University Press.
- Bauer, Laurie. 2001. *Morphological productivity* (Cambridge Studies in Linguistics 95). Cambridge: Cambridge University Press.
- Benczes, Réka. 2006. *Creative compounding in English: The semantics of metaphorical and metonymical noun-noun combinations* (Human Cognitive Processing 19). Amsterdam & Philadelphia: John Benjamins.
- Benczes, Réka. 2010. Setting limits on creativity in the production and use of metaphorical and metonymical compounds. In Sascha Michel & Alexander Onysko (eds.), *Cognitive approaches to word formation*, 221–245. Berlin & New York: Mouton de Gruyter.
- Berry, Roger. 2005. Making the most of metalanguage. *Language Awareness* 14(1). 3–20.

- Bower, Jack & Satomi Kawaguchi. 2011. Negotiation of meaning and corrective feedback in Japanese/English eTandem. *Language Learning & Technology* 15(1). 41–71.
- Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form* (Typological Studies in Language 9). Amsterdam & Philadelphia: John Benjamins.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Fischer, Roswitha. 1998. *Lexical change in present-day English: A corpus-based study of the motivation, institutionalization, and productivity of creative neologisms*. Tübingen: Narr.
- Kerremans, Daphné. 2015. *A web of new words: A corpus-based study of the conventionalization process of English neologisms*. Frankfurt am Main etc.: Peter Lang.
- Metcalf, Allan. 2002. *Predicting new words*. Boston: Houghton Mifflin.
- Morzycki, Marcin. 2011. Metalinguistic comparison in an alternative semantics for imprecision. *Natural Language Semantics* 19. 39–86.
- Predelli, Stefano. 2003. Scare quotes and their relation to other semantic issues. *Linguistics and Philosophy* 26. 1–28.
- Schmid, Hans-Jörg. 2008. New words in the mind: Concept-formation and entrenchment of neologisms. *Anglia* 126 (1). 1–36.
- Schmid, Hans-Jörg. 2016. *English morphology and word-formation: An introduction*. 3rd revised and extended edition. Berlin: Erich Schmidt.
- Semino, Elena. 2016. A corpus-based study of “mixed metaphor” as a metalinguistic comment. In Raymond W. Gibbs Jr. (ed.), *Mixing metaphor* (Metaphor in Language, Cognition, and Communication 6), 203–220. Amsterdam: John Benjamins.
- Svanlund, Jan. 2006. Hur man etablerar sig som *curlingförälder* och *stafettläkare*. *Svenskans beskrivning* 28. Örebro universitet. 359–368.
- Svanlund, Jan. 2009. *Lexikal etablering: En korpusundersökning av hur nya sammansättningar konventionaliseras och får sin betydelse* (Stockholm Studies in Scandinavian Philology, n.s., 52). Vällingby: Elanders.