

Nominal word formations and syntagms with *super*. A corpus-based investigation

Introduction

This article presents a corpus-based investigation of the factors that co-vary in contemporary written German with the use of nominal word formations and syntagms construed with *super*. Over the course of the last decades, the usage frequency of *super* as a bound morpheme (*Super-N*) and unbound morpheme (*super + N*) has significantly increased; cf. *Superteam*, *Superstimmung*, *Superinfektion*, *Superrechner* and *super Team*, *super Stimmung*, *super Angelegenheit*, *super Pünktlichkeit* etc. The morpheme *super* has been repeatedly described in the past (see Wendelken 1967; Kann 1973; Reinhardt 1975; Schulz et al. 1978; Schmidt 2005; Ruf 1996; Lohde 2006), the alternation between the two constructions, however, has not yet been systematically investigated (cf. Kammerer 2001, Grzega 2004 and Fleischer/Barz 2012 who discuss, among other things, the various accentuations of the alternations). This paper focuses on whether there are factors that have a statistically significant correlation for a preference towards one of the two constructions, and which conclusions about functional differences can therefore be drawn. We hypothesise that alternation can be semantically functional and is more than a mere orthographical issue. The aim of this paper is to close the existing empirical gap in academic literature with the help of a corpus-based analysis in which a number of quantifiable factors are distinguished that are potentially relevant for the alternation between the two factors.

Method

A representative sample of 1000 occurrences was compiled on the basis of the *Deutsches Referenzkorpus* (DEREKO, *W-Archiv der geschriebenen Sprache*; www1.ids-mannheim.de/kl/projekte/korpora.html). A total of 500 word formations (*Super-N*) and 500 syntagms (*super-N*) were chosen at random. The examples were collected exclusively from sub-corpora that comprise standard German texts. Austrian and Swiss sources, as well as the Wikipedia corpus were excluded from the data collection. The occurrences were annotated and analysed with regard to six variables that capture the semantic, morphosyntactic, text-linguistic and etymological features. The six variables are: classification of semantic class (following Ruf 1996), additional attributive modification (the NP either includes an additional attribute, as in *der neue Superflughafen* and *Bernie Ecclestones Superevent* or not, as in *eine Superinfektion*), nature of reference (concrete, e.g., *Super-Torhüter Freddy Brathwaite*, or abstract, e.g., *super Stimmung*), determination (definite, e.g., *die Super-Zicke*, or indefinite, e.g., *ein super Wochenende für Superreiche*), length of text (short text, e.g., title, lead text, heading or sub-heading etc., or longer texts) and etymology of the base word (foreign base, e.g., *ein Super-Comeback*, or German base, e.g., *ein super Abend*). The quantitative differences are interpreted with the help of a multivariate analysis (classification tree method).

All occurrences are assigned to one of the semantic classes according to Ruf (1996). This distinction aims to categorise both the word formations as well as the syntagms into groups of conventional usage in order to identify and compare the different referencing functions of *super* in both constructions. In this respect, we agree with Schmidt (2005, p. 415) who says that if the two constructions are not distinguished, we would „auf etwas verzichten, was zu den Forderungen der Sprachkultur gezählt werden muss, nämlich auf eine optimale Ausnutzung der unterschiedlichen Möglichkeiten sprachlicher Mittel, zu der auch die differenzierte Verwendung lexikalischer Einheiten gehört“.

Schmidt (2005) and Ruf (1996) are both concerned with the semantics of *Super-*. According to Schmidt, it is possible to differentiate two “basic meanings” of *Super-*. On the one hand, the morpheme conveys that an object or a person „sich über bzw. auf einer anderen [Sache oder Person] befindet, was konkret räumlich, in einer Skala, in einer Hierarchie oder in Form einer beaufsichtigenden Überordnung der Fall sein kann“ (Schmidt 2005, p. 409). On the other, *Super-* means „daß in der Kombinationsbedeutung ein Hinausgehen über etwas erscheint, und zwar ein Hinausgehen über etwas, das allgemein als normal oder normal erwartbar angesehen wird“ (ibid., p. 410). As a result, these two basic meanings are divided into further sub-categories (see ibid., p. 410–413). On the whole, Ruf (1996) also applies Schmidt’s (2005) classification. In her analysis, however, she merges two of Schmidt’s classes into one single class and adds two further categories.

It should be borne in mind that the nine “semantic classes” defined by Ruf (1996, p. 113–121), and that we refer to in the annotation of our samples, do not distinguish nine different meanings of the morpheme *super* but instead represent semantic classes that are achieved through the combination with a base word or noun. It would obviously be circular to claim that the prefix has a positive meaning in occurrences such as *Supersieg*, *Superkünstler*, *Supertreffer*, *Super-Chefarzt*, *super Tempospiel*, *super Trainer*, *super Reflex* and a negative meaning in *Superschurke*, *Superterrorist*, *Superkatastrophe*, *Super-Fiesling* and *Super-Lügner*. In reality, the meaning of *super/Super-* is the same in both classes. In the one case, however, the semantics of the respective base words or nouns triggers – provided that no other negating cues exist in the sentence – a positive interpretation (*Sieg*, *Künstler*, *Chefarzt*, *Trainer* etc.) while it prompts a negative interpretation in the other (*Schurke*, *Terrorist*, *Katastrophe* etc.). The distinguished “semantic classes” therefore represent classes of conventional or “norm” usage of the two constructions *Super-N* and *super + N* (see Willems 2001 on the concept of norms in the analysis of word formations). The interpretation of an occurrence in actual discourse can, however, vary depending on context. Context is understood as „sowohl Satz und Text als auch die Situation“ (Biedermann 1969, p. 90, quoted in Ruf 1996, p. 70).

Quantitative analysis

We apply the so-called classification tree method (see Baayen 2008, p. 148–154), type “conditional inference tree” (CTREE) for the multivariate analysis. The analysis indicates that four variables – distinction by semantic class, determination, concreteness of reference, and attributive modification – correlate significantly with the alternation between the word formation construction *Super-N* and the syntagm *super + N*. In this instance, the classification of semantic classes plays the most important role. The second class in particular (‘Das in der Basis Genannte geht über das, was allgemein als normal oder normal erwartbar angesehen wird, hinaus’) is the most influential: five of the six terminal nodes

in the classification tree include this factor, which is the equivalent of more than three quarters of all occurrences. Determination also constitutes an important factor. According to the analysis, the variables length of text and etymology of the base word do not influence alternation.

On a whole, the statistic model's predictive value is relatively high: 79% of occurrences are classified correctly. In regard to word formation (88%), however, the model achieves much better results than for syntagms (74%). This suggests that the interaction of the various variables is better at capturing the word formations than the syntagms in the dataset and that the occurrence of word formations has a more systematic relationship with the variables than the occurrence of syntagms.

Qualitative analysis

Between them, the significant factors of the multivariate analysis document a function of the word formation construction *Super-N* that can, following Willems (1990, p. 64–66, 72) and in contrast to the construction *super + N*, be considered a “proprial” function. This refers to a form of semantic specialisation that labels numerous word formations (see the discussion in Brugmann 1900 and Paul 1903, and the “labeling” function in Kastovsky 1986). The “proprial” function consists of the word formation providing a more concise designation for an object or situation than the respective syntagm, thereby identifying it as a cognitive entity and making it available for discourse. This type of function can occasionally be found in the corpus itself, e.g. in the article “Vom super Rechner zum Superrechner” from the Braunschweiger Zeitung of 05 July 2007:

Es gibt einen Unterschied zwischen einem super Rechner und einem Superrechner. Ein super Rechner ist derjenige, der dieser Kolumne den Namen gegeben hat. Als der kleine Carl neun Jahre alt ist, haut es seinen Mathelehrer um, so schnell löst er die Aufgaben. [...] Ein Superrechner ist ein Hochleistungsrechner, der innerhalb kurzer Zeit unglaublich viele Rechenoperationen ausführen kann.

The use of a definite determiner, additional attributive modification with an attributive adjective or genitive as well references to a concrete noun significantly correlate with the use of the word formation construction *Super-N*. By contrast, indefinite or lacking determination as well as lacking attributive modification correlate with the use of the syntagm *super + N*. Finally, there is no significant preference for one of the two constructions in the case of references to concrete nouns. The use of definite determiners, additional attributive modifications and concreteness of references profile the proprial function of the word formation construction. It is distinguished from the respective syntagm by a formal unit (univerbation) that makes it appear particularly suitable for the identification of a person, object or situation. This observation is consistent with the iconic rationale of the correlation between the use of word formations and references to the concrete. Furthermore, the sample exhibits numerous instances of full-fledged proper nouns, which are co-referential with word formations, being used in the immediate context of word formations of the type *Super-N* (either immediately following or in the sentence before or after), for example *das deutsche Super-Model Heidi Klum, der amerikanische Super-Magier David Copperfield, das Superflugzeug A380, der neue Superhartstoff Osmium, Superkomet Hale-Bopp, Superpapst Karol Wojtyla, die Superweltmacht China* etc. This type of co-occurrence is less frequent in the case of syntagms. It is also interesting to note that word formations are frequently combined with a definite determiner, which is also comparable to the inherent individualisation of proper nouns that are normally not reliant on a determiner (see Wil-

lems 1996, p. 107f.). The frequent use of definite determiners and additional attributes underlines a feature of word formation constructions that highlight their contrast to syntagms at both the discourse level (in a text) and the cognitive level (as a process with which to identify the denotatum). In some cases, the word formation itself can assume the status of a proper noun, resulting in a common noun's designating ("proprial") function yielding to the, in the strictest sense, "onymic" function (proper noun), e.g., *Superbenzin* and *Superphosphat*.

Final remarks

All specified factors - determination, attributive modification, concreteness and the broad distribution across multiple semantic classes - naturally constitute tendencies (that have a statistically significant correlation with the word formation constructions) and not exclusion criteria (that allow for a rigid separation between the word formation construction *Super-N* and the syntagm *super + N*). This is also entirely in the spirit of the multivariate analysis of the alternation explored in this paper, which shows that the use of two possible constructions (*Super-N* and *super + N*) is associated with multiple factors and does not have a monocausal explanation. It is possible that, in individual cases, the choice between *Super-N* and *super + N* is merely an orthographical decision. However, the results of the statistical analysis show that different functional factors exist that correlate with the alternation to varying degrees.

The question regarding the extent to which the interaction of the observed relevant factors reflects a psychological reality cannot be resolved in this paper. In any event, a distinction must be made between the status of a linguistic method and the (not merely psychological but also cultural and social) reality of language itself. This distinction has gained more and more traction over the course of the last two decades, ever since the "quantitative turn" in linguistic methodology. It is entirely possible that, for an individual speaker, a single factor might determine which of the two alternations is the most suitable in a specific discourse without the other factors playing a comparable role. In other cases, it might be precisely the interaction between the factors that (co)determines the choice (see the hypothesis on "quantitative harmonic alignment", Bresnan/Ford 2010, p. 181). However, these decisions affect individual discourse activity that cannot be captured with the help of quantitative analytical methods such as classification trees. Instead, these types of methods are based on sufficiently large datasets that make it possible to ignore individual differences in favour of generalising typifications in order to uncover tendencies in the products of language use. On these grounds, it would also be highly interesting to compare the results of this analysis with the results of an informant survey in which native speakers rate sentences that consider the factors analysed here and potentially further modify them. By doing so, it would be possible to investigate to what extent individual preferences towards one of the two alternations actually coincide with the results generated here, and in what way the interaction between various factors plays a role in individual language production.

Comparing written and spoken language would most likely prove to be difficult as an analysis of the postulated prosodic differences between both alternations in spoken language would hardly be feasible. It would make sense, however, to conduct further studies based on corpus analyses that also take into account other factors such as, e.g., the information-structural status of both constructions in a sentence or text (cf. the observations

that word formations frequently occur with co-referential proper nouns) and differentiation according to text type and register. It would also be fruitful to include the time axis to investigate whether identified tendencies in written language are historically stable or variable. The pejorative function of *Super-*, discussed in older studies, could also be an interesting subject for a diachronic investigation.

References

- Baayen, Harald (2008): Analyzing linguistic data: A practical introduction to statistics using R. Cambridge: Cambridge University Press.
- Bresnan, Joan/Ford, Marilyn (2010): Predicting syntax: Processing dative constructions in American and Australian varieties of English. In: *Language* 86, p. 168–213.
- Brugmann, Karl (1900): Über das Wesen der sogenannten Wortzusammensetzungen. Eine sprachpsychologische Studie. In: *Berichte über die Verhandlungen der königlich sächsischen Gesellschaft der Wissenschaften zu Leipzig. Philologisch-historische Classe* 52, p. 359–401. [Wieder abgedruckt in: Lipka, Leonhard/Günther, Hartmut (Hg.) (1981): *Wortbildung*. Darmstadt: Wissenschaftliche Buchgesellschaft. p. 135–178.]
- Fleischer, Wolfgang/Barz, Irmhild (2012): *Wortbildung der deutschen Gegenwartssprache*. 4., völlig neu bearb. Aufl. Tübingen: De Gruyter.
- Grzega, Joachim (2004): Ein Spitzenpolitiker ist nicht immer ein Spitzen-Politiker. Wie man prosodische Akzente nutzt, um semantische ‚Akzente‘ zu setzen. In: *Muttersprache* 114, p. 321–332.
- Kammerer, Matthias (2001): Verstärkungsbildungen im Deutschen. Versuch einer phänomenologischen Bestimmung. In: Lehr, Andrea (Hg.): *Sprache im Alltag. Beiträge zu neuen Perspektiven in der Linguistik*. Berlin: De Gruyter. p. 293–319.
- Kann, Hans-Joachim (1973): Spielfreude in der Sprache: ‚Super-‘ und ‚Mini-‘. In: *Muttersprache* 83, p. 198–210.
- Kastovsky, Dieter (1986): The problem of productivity in word formation. In: *Linguistics* 24, p. 585–600.
- Lohde, Michael (2006): *Wortbildung des modernen Deutschen. Ein Lehr- und Übungsbuch*. Tübingen: Narr.
- Reinhardt, Werner/Frey, Christa (1975): *Deutsche Fachsprache der Technik: Ratgeber für die Sprachpraxis*. Leipzig: Verlag Enzyklopädie.
- Ruf, Birgit (1996): *Augmentativbildungen mit Lehnpräfixen. Eine Untersuchung zur Wortbildung der deutschen Gegenwartssprache*. Heidelberg: Winter.
- Schmidt, Günter Dietrich (2005 [1990]): Super- und top-. Ein Vergleich von zwei im Deutschen heute produktiven Lehnpräfixen. In: Müller, Peter O. (Hg.): *Fremdwortbildung. Theorie und Praxis in Geschichte und Gegenwart*. Frankfurt: Peter Lang. p. 407–415. [Ersterscheinung 1990 in: *Muttersprache* 100, p. 204–210.]
- Schulz, Hans/Basler, Otto/Strauß, Gerhard (1978): *Deutsches Fremdwörterbuch*. 2. Aufl. Berlin: De Gruyter.
- Wendelken, Peter (1967): Der Einfluß des Englischen auf das heutige Werbedeutsch. In: *Muttersprache* 77, p. 289–308.
- Willems, Klaas (1990): Tageshöchsttemperaturen, Billigst-Flüge und Halbknaben. Zur Syntax, Semantik und Stilistik eines beliebten Wortbildungsmodells im heutigen Deutsch. In: *Deutsche Sprache* 18, p. 52–75.