Are *doggies* cuter than *dogs*? Emotional valence and concreteness in German derivational morphology

Gabriella Lapesa, Sebastian Padó, Tillmann Pross, and Antje Roßdeutscher
University of Stuttgart, Institute for Natural Language Processing
[gabriella.lapesa,pado,tillmann.pross,
antje.rossdeutscher]@ims.uni-stuttgart.de

Abstract

The semantic behavior of derivational processes has been investigated with compositional distributional models relating the meaning of base, affix, and derivative (e.g., *anti+capitalist* → *anticapitalist*). While broadly successful, these approaches model how the distributional behavior generally is affected by derivation. Meanwhile, their predictions can not be interpreted at the level of linguistic regularities. In this paper, we adopt an alternative approach and focus on the impact of derivation on finer-grained semantic properties of the base. We focus on (the psycholinguistically prominent) *emotional valence*, i.e., the speakers' positive/negative evaluation of the word referent. We present two case studies on German derivational patterns, combining distributional and regression analysis. We are able to establish the broad presence of valence effects in German derivation as well as strong interactions with concreteness.

1 Introduction

Morphological derivation (Plag, 2003) is a word formation process which combines *bases* (e.g., *Hund* – "dog") with *affixes* (e.g., the diminutive *-chen*) into new words (*Hündchen* "doggie"). The semantic properties of derivation have been extensively explored in theoretical linguistics, and a number of recent computational studies in compositional distributional semantics have modelled the mappings that hold between the vectors of bases, affixes, and derivatives (Lazaridou et al., 2013; Luong et al., 2013; Padó et al., 2016). What these studies crucially lack, though, is *interpretability*: typically, they model mappings in an embedding space, but have little to say about linguistic regularities such as systematic changes in *meaning components*.

In this paper, our goal is to do exactly that, namely investigate the effects of derivation on a specific meaning component, *emotional valence* (henceforth, valence), which quantifies the speaker's positive or negative affect towards the referent of a word. This choice is motivated by psycholinguistic considerations: Valence is very well established in the literature as having substantial effects on human language processing (Vinson et al., 2014; Kuperman et al., 2014; Snefjella and Kuperman, 2016). Since it is not clear that the effects on valence take place independently of other variables, we extend our analysis to include a set of other meaning components, most notably *concreteness*, a second prominent meaning component in psycholinguistics (cf. the references above). Both meaning components are also highly relevant for NLP: (Variants of) valence occur under the names of sentiment and polarity and form the basic variable of interest in sentiment analysis (Pang and Lee, 2008). Concreteness is exploited, among other things, for metaphor identification (Turney et al., 2011; Köper and Schulte im Walde, 2016b).

The questions we ask are (a) whether a derivative carries a significantly different valence from its base; and (b) whether there are interactions between valence and concreteness (i.e., whether valence shifts occur only in more concrete vs. abstract contexts). To the best of our knowledge, the effect of derivation on valence has not been explored in distributional semantics. Our work extends a couple of studies that consider the interaction between valence and concreteness: Mohammad et al. (2016) present a collection of ratings targeting emotion and metaphor; Hill and Korhonen (2014) explore the interplay between subjectivity and concreteness. From a purely linguistic perspective, valence is situated between semantics and pragmatics; despite the interest for the interplay between semantics and pragmatics in derivational morphology (Dressler and Barbaresi, 1998; Plag, 2003), there has been no attempt yet to integrate theoretical considerations and computational modeling.

Our contribution is twofold. First, computationally, we define a distributional procedure that quantifies the basis–derivative differences with respect to specific meaning components and aggregates these differences across a large vocabulary with a regression analysis. Second, linguistically, we present two case studies on German derivation. The first one focusses on a specific pattern (*über- (over-)* prefix verbs) and illustrates the integration of valence with a theoretically motivated manual subclass analysis. The second one targets a larger set of patterns without manual annotation. We establish a strong presence of valence effects in derivation, even where its role would have been not obvious (female forms are used in more positive contexts than their male counterparts). We also find an interaction with concreteness which characterizes a "classic" evaluative pattern, the diminutive *-chen* (see Jurafsky (1996) for a cross-linguistic overview of the semantic spectrum covered by diminutives), as well as a pattern which has a clear evaluative flavor, the adversative *anti-*.

2 Experimental Setup

Our goal is to analyse the role of *valence* as a meaning component that undergoes systematic changes between base and derived words. We proceed as follows: given a set of base-derived word pairs for a derivation pattern, we represent the words distributionally. Then we quantify valence and a set of auxiliary meaning components (concreteness, imageability, arousal) for each word using a lexicon-based approach. Finally, we perform a regression analysis to analyse the factors affecting valence.

Distributional Semantic Model Since our assignment of meaning components is lexicon-based (see next paragraph), we require a distributional model with lexical dimensions. This precludes the use of neural embedding models (Mikolov et al., 2013). Instead, we use a count (bag-of-words) distributional model with lexical dimensions. It is extracted from SdeWaC (Faaß and Eckart, 2013), a 800M words German web corpus with a large target and context vocabulary (approx. 280k lemmatized open-class words). We adopt standard choices for the main parameters, namely a symmetrical 5-words context window and positive pointwise mutual information to transform raw counts.

Computing Meaning Components. We employ the German Affective Norms (Köper and Schulte im Walde, 2016a). The dataset contains automatically generated scores for 350k German lemmas on a 0 to 10 scale for four psycholinguistically prominent meaning components: *valence*, the (un-) pleasantness associated with the word; *arousal*, the intensity of the emotion associated with it; *concreteness*, the extent to which the word's referent can be perceived; and *imageability*, the extent to which the word's referent can be perceived visually. While the scores on these four components can in principle be used 'as is' for words covered by the resource, we found that their quality can be crucially improved (see Section 3 for details) by defining a *context-based reweighing scheme*. We define the score assigned to a target word t on a given component as the weighted average of the scores of its context words c by computing the dot product between the (L1-normalized) distributional vector for t and the vector of Affective Norm Scores for all context words. For each component, we reduce the set of context words with scores belonging to the top and bottom quartile for this component.

Regression analysis. To gain a systematic understanding of valence, we perform a linear regression analysis. Linear regression predicts a continuous dependent variable (here, the valence score) as a linear combination of weighted predictors. We considered (a), theoretically motivated subclasses of *über-* verbs (Study 1, Section 3) and derivational patterns (Study 2, Section 4); (b), the meaning dimensions annotated in the German Affective Norms (imageability, concreteness, arousal); (c), frequency effects, as is best practice. In a model selection step, we discarded imageability based on a collinearity analysis (strong correlation to concreteness) and added the interaction between class/pattern and the Affective scores that were significant for both studies. The final model is:[1]

$$\begin{aligned} &\texttt{valence} \sim \texttt{class/pattern} \\ &\star\ \texttt{(concreteness + arousal)} \\ &+\ \texttt{freq\_base + freq\_derived} \end{aligned} \quad (1)$$

We trained three regression models to predict valence scores for base and derived words and to predict *differences* between valence scores for base and derived words.

3 Study 1: *über* prefix verbs

We investigate *über-* prefix verbs as an interesting object in lexical semantics: some *über* verbs (e.g., *überrennen*, "to overrun", *überschwemmen*, "to overflood",) encode a negative evaluation for events perceived as uncontrolled or uncontrollable (an excess reading, absent in the corresponding base terms *rennen*, "to run", *schwemmen*, "to float"). We build on a previous study of *über* prefix verbs (Pross and Roßdeutscher, 2015) that has produced a dataset of 74 *über* verbs and their corresponding bases manually selected to ensure that derived words are transparent with respect to their bases, at least in their dominant reading. Each pair was manually assigned to one of four theoretically motivated classes that differ by the contribution of *über-* to the interpretation of prefix verb:

- TRANSFER of an object from a source region to a goal region (16 pairs). Ex: *bringen, überbringen* ("to bring", "to deliver").
- APPLICATION of an object to another object (19 pairs). Ex: *kleben, überkleben* ("to paste", "to paste over").
- movement ACROSS some boundary or obstacle, which is conceptualized as a patient and in some cases undergoes change of state (18 pairs). Ex: *fahren, überfahren* ("to drive", "to drive (something) over").
- exceeding a certain threshold on a scale (MORE) provided by the base verb or by the usage context (21 pairs). Ex: *(be)werten, überbewerten* ("to value", "to overvalue").

We hypothesize that the ACROSS class is associated with negative valence, the others are neutral. We test the hypothesis by including the class in our regression model (cf. Equation (1)).

Results. Table 1 summarizes the fit of the linear models in terms of their ability to explain the valences of base verbs (column "Base"), the valences of *über* prefix verbs (column "Derived"), and the differences between base and prefix verbs (column "Shift"). It shows both the total amount of variance accounted for and the contribution of individuals predictors, computed though Lindeman-Merenda-Gold (LMG) scores (Lindeman et al., 1980). The fit of the full models (between .60 and .74 adjusted R^2) is very good, and even though frequencies are a major predictor (as almost always), both semantic classes and other meaning components (concreteness, arousal) contribute nicely.

Table 2 shows coefficients for all predictors that are significant in at least one of the columns. In the following, we focus on the Shift results and give Base and Derived for comparison only. For Shift, a positive coefficient for a predictor means that derived words with a high value of the predictor exhibit a higher valence than their bases. Vice versa, a predictor with a negative coefficient will reduce the

[1] We use the R statistical environment. The asterisk in the formula represents the interaction between class/pattern and concreteness and arousal. Continuous predictors are scaled, categorical variables are sum-coded: effects are calculated with the grand mean of the groups as reference value. Frequencies are log-transformed.

Predictor	Shift	Derived	Base
Semantic Class	.031	.058	.072
Concreteness	.037	.051	.087
Arousal	.011	.002	.108
Class:Arousal	.088	.086	.004
Base frequency	.105	N/A	.498
Derived frequency	.384	.553	N/A
Adjusted R^2	.60 ***	.72 ***	.74 ***

Table 1: Study 1 model fit (explained variance)

Predictor	Shift	Derived	Base
APPLICATION	–	–	-.06 **
Concreteness	-.06 **	-.08 **	–
ACROSS:Arousal	-.13 ***	-.10 ***	–
Base frequency	.15 ***	N/A	-.15 ***
Derived frequency	-.20 ***	-.20 ***	N/A

Table 2: Study 1: Coefficients of predictors

valence scores of derived words associated with high values of it. Contrary to our expectations, there is no significant main effect for any semantic class in the Shift analysis, meaning that the verb classes at large do not differ in valence. We do however, specifically find an interaction between the ACROSS semantic class and arousal that is highly significant and has a negative sign. Thus, ACROSS bases do tend to acquire negative valence as you add *über-*, but only if they already carry high arousal, i.e., are "emotionally loaded" verbs. A second interesting observation is the negative main effect of concreteness. It shows that across all pairs in the dataset, negative valence shifts are more pronounced for concrete verbs (*fahren, überfahren* "drive, drive over") than for abstract verbs (*nehmen, übernehmen* "take, take over").

Finally, we return to a question from §2: is there a difference between using valence scores from the Affective Norms and (re-)computing them distributionally? We repeated the analysis above using the Affective Norms valence scores, and found a much lower model fit (only .21 adjusted R^2, compared to .60 as in Table 1) as well as an absence of significant effects. In sum, the distributional valence scores do a substantially better job.

4 Study 2: Other Derivation Patterns

Our second study extends the focus beyond *über-* to six other German within part-of-speech derivation patterns from a previous study (Kisselew et al., 2015):

- N→N, FEMALE: *-in* (80 pairs). Ex: *Bäcker, Bäckerin* ("baker", "female baker")
- N→N, DIMINUTIVE: *-chen* (80 pairs). Ex: *Schiff, Schiffchen* ("ship", "small ship")
- A→A, OPPOSED: *anti-* (80 pairs). Ex: *religiös*, *antireligiös* ("religious", "antireligious")
- A→A, NEGATIVE: *un-* (80 pairs). Ex: *dankbar, undankbar* ("grateful", "ungrateful")
- V→V, DIRECTED: *an-* (68 pairs). Ex: *sprechen, ansprechen* ("to speak", "to address")
- V→V, TRAVERSE: *durch-* (70 pairs). Ex: *gehen, durchgehen* ("to go", "to go through")

Here, our hypotheses are that DIMINUTIVE comes with a positive valence shift and ADVERSE and DIRECTIONAL with a negative valence shift.

Predictor	**Shift**	**Derived**	**Base**
Pattern	.082	.093	.076
Concreteness	.009	.001	.030
Arousal	.002	.006	.001
Pattern:Concreteness	.018	.009	.008
Base frequency	.135	N/A	.524
Derived frequency	.148	.277	N/A
Adjusted R^2	.38***	.37***	.63***

Table 3: Study 2 model fit (explained variance)

Predictor	**Shift**	**Derived**	**Base**
AN-	-.06 *	–	.04 ***
ANTI-	–	-.05 *	-.03 *
-IN	.06 *	.07 **	–
Concreteness	-.04 ***	–	-.01*
ANTI-:Concreteness	.07 **	.06 *	–
-CHEN:Concreteness	.04 *	–	–
Base frequency	.16 ***	N/A	-.16 ***
Derived frequency	-.17 ***	-.18 ***	N/A

Table 4: Study 2: Coefficients of predictors

Results. We again start with model fit (Table 4). While the fit is lower than in Study 1, the new dataset is much more varied. Thus, we consider the (highly significant) Adjusted R^2 of .38 as still very good. Again, frequency explains much of the variance, followed by the derivational pattern.

The coefficients in Table 4 again show that our hypotheses hold up only partially. A significant negative effect for the AN- pattern is explained by the corresponding results for the base verbs, which show a highly significant positive valence compared to all other patterns in the dataset. We do not find a main effect for -CHEN, but a positive interaction with concreteness: concrete objects (*Hund*, "dog") gain in valence through diminution (*Hündchen*, "doggie") while abstract objects do not or can even acquire a pejorative component (*Idee, Ideechen* "idea, little idea"). A comparable interpretation offers itself for ANTI- where again there is no main effect but an interaction with concreteness (compare the strongly negative *antisemitisch* with the neutral *antibiotisch*). A somewhat unexpected result is the positive main effect of the female pattern -IN. A possible interpretation is that the marked female forms, many of which are professions, are only chosen when the gender is relevant, which is supported by the occurrence of positive evaluative adjectives ("good", "skilled"). In this connection, our results are a contribution to the characterization of gender bias in language (cf., Terkik et al. (2016) for another example of such study). At any rate, a more detailed analysis of the contexts is required to understand this effect better. Lack of a negative effect for *über* shows that for certain patterns an approach which is based on semantic subclasses of the derived terms is necessary to detect valence shifts that are more fine-grained.

As in Study 1, there was an overall negative effect of concreteness; however, this time, arousal did not play any significant role (its interaction being specific to the semantic classes annotated in the *über* dataset).

5 Conclusion

In this study, we have applied a kind of "magnifying glass" approach: instead of attempting to characterize the meaning of a word as completely as possible from distributional evidence, we focus on a small set of specific meaning components centered around emotional valence, and investigated how the strength of these components is influenced by derivational word formation. We described a method that can be used to extract a data-driven analysis of valence shifts and their interactions with other variables: It maps distributional representations for the words onto a valence scale and uses regression analysis as a pattern mining framework. We showed that the method can use manual annotation when available (Study 1) but also scales to larger, automatically generated datasets (Study 2). Beyond valence, our approach is applicable to other meaning components. Our analysis has uncovered a number of novel observations, notably the modulation of emotional valence for prefix verbs encoding boundary crossing (Study 1) and the unexpected presence of a positive evaluative meaning nuance in the female pattern (Study 2), as well as interactions between factors (Study 1 and Study 2). In particular, we found a strong effect of concreteness in modulating emotional valence shifts in derivation.

Acknowledgments

We gratefully acknowledge funding of our research by the DFG, SFB 732 (project B9: Lapesa and Padó; project B4: Pross and Roßdeutscher).

References

Dressler, W. U. and L. M. Barbaresi (1998). Morphopragmatics. In J. B. Jef Verschueren, Jan-Olaf Östman and C. Bulcaen (Eds.), *Handbook of Pragmatics: 1997 Installment*. John Benjamins Publishing.

Faaß, G. and K. Eckart (2013). Sdewac – a corpus of parsable sentences from the web. In *Language Processing and Knowledge in the Web*, Volume 8105 of *Lecture Notes in Computer Science*, pp. 61–68. Springer Berlin Heidelberg.

Hill, F. and A. Korhonen (2014). Concreteness and subjectivity as dimensions of lexical meaning. In *Proceedings of ACL*, Baltimore, USA, pp. 725–731.

Jurafsky, D. (1996). Universal tendencies in the semantics of the diminutive. *Language 72*(3), 533–578.

Kisselew, M., S. Padó, A. Palmer, and J. Šnajder (2015). Obtaining a better understanding of distributional models of German derivational morphology. In *Proceedings of IWCS*, London, UK, pp. 58–63.

Köper, M. and S. Schulte im Walde (2016a). Automatically generated affective norms of abstractness, arousal, imageability and valence for 350000 German lemmas. In *Proceedings of LREC*, Portoroz, Slovenia, pp. 2595–2598.

Köper, M. and S. Schulte im Walde (2016b). Distinguishing literal and non-literal usage of German particle verbs. In *Proceedings of NAACL-HLT*, San Diego, USA, pp. 353–362.

Kuperman, V., Z. Estes, M. Brysbaert, and A. B. Warriner (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology. General 143*(3), 1065–1081.

Lazaridou, A., M. Marelli, R. Zamparelli, and M. Baroni (2013). Compositional-ly derived representations of morphologically complex words in distributional semantics. In *Proceedings of ACL*, Sofia, Bulgaria, pp. 1517–1526.

Lindeman, R. H., P. F. Merenda, and R. Z. Gold (1980). *Introduction to Bivariate and Multivariate Analysis*. Glenview, IL, USA: Scott Foresman.

Luong, M.-T., R. Socher, and C. D. Manning (2013). Better word representations with recursive neural networks for morphology. In *Proceedings of CoNLL*, Sofia, Bulgaria, pp. 104–113.

Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013). Efficient estimation of word representations in vector space. In *Proceedings of ICLR*, Scottsdale, AZ.

Mohammad, S. M., E. Shutova, and P. D. Turney (2016). Metaphor as a medium for emotion: An empirical study. In *Proceedings of STARSEM*, Berlin, Germany, pp. 23–33.

Padó, S., A. Herbelot, M. Kisselew, and J. Šnajder (2016). Predictability of distributional semantics in derivational word formation. In *Proceedings of COLING*, Osaka, Japan, pp. 1285–1296.

Pang, B. and L. Lee (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval 2*(1–2), 1–135.

Plag, I. (2003). *Word-formation in English*. Cambridge Textbooks in Linguistics. Cambridge University Press.

Pross, T. and A. Roßdeutscher (2015). Measuring out the relation between conceptual structures and truth-conditional semantics. In *Selected papers from the Workshop "Bridging Formal and Conceptual Semantics"*, Düsseldorf.

Snefjella, B. and V. Kuperman (2016). It's all in the delivery: Effects of context valence, arousal, and concreteness on visual word processing. *Cognition 156*, 135–146.

Terkik, A., E. Prud'hommeaux, C. O. Alm, C. Homan, and S. Franklin (2016). Analyzing gender bias in student evaluations. In *Proceedings of COLING*, Osaka, Japan, pp. 868–876.

Turney, P. D., Y. Neuman, D. Assaf, and Y. Cohen (2011). Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of EMNLP*, Edinburgh, United Kingdom, pp. 680–690.

Vinson, D., M. Ponari, and G. Vigliocco (2014). How does emotional content affect lexical processing? *Cognition and Emotion 28*(4), 737–746.