

An Inquiry into the Semantic Transparency and Productivity of German Particle Verbs and Derivational Affixation

Inna V. Stupak & R. Harald Baayen
University of Tübingen

October 5, 2022

Abstract

This study addresses the relation between morphological productivity and semantic transparency. Using distributional semantics, we compare German word formation with particles with derivational word formation. We observed that derivational suffixes, but not particles, tend to make strong independent semantic contributions to their carrier words. In two-dimensional t-SNE maps, complex words show clustering by affix, but not by particle. Furthermore, the semantic vectors of suffixed words are predictable from their base words with higher accuracy than is possible for particle verbs. This holds irrespective of whether the semantic vectors of complex words are predicted using a generalization of the FRACCS model, or whether a more recent, and more precise, additive model (CosClassAvg) is used. For particle verbs, but not affixed verbs, a greater semantic similarity within the set of complex words correlated negatively with the number of types. Only for particle verbs, a greater number of types predicted a reduced probability of unseen types. Considered jointly, the particle verbs are characterized by quantitative properties that are very similar to those observed for Mandarin compounds. We propose that particle verbs, just as compounds, primarily serve the onomasiological function of labeling, resulting in highly distinct semantic vectors for particle verbs. By contrast, derivational affixation groups complex words closely together in semantic space while maintaining strong and consistent semantic relations with the base words. This enables these words to serve not only as labels, but also allows them to be used for syntactic recategorization.

1 Introduction

For the quantitative evaluation of morphological productivity, two measures have been found especially useful. The first measure specifies the number of word types (or vocabulary size) $V(N)$ that occur in a sample of N tokens.¹ The second measure presents the rate at which the number of types (vocabulary size) is increasing. This growth rate of the vocabulary size can be estimated by the ratio of words occurring once only ($V(1, N)$) and the sample size N : $\mathcal{P}(N) = V(1, N)/N$ (see, e.g., Good, 1953; Baayen, 2001). The two measures capture different aspects of productivity: $V(N)$ highlights the extent to which a morphological category is in use, whereas $\mathcal{P}(N)$ estimates the probability that when the sample size is increased, previously unseen types will be observed. Research on the productivity of English affixes has shown that for this language, $V(N)$ and $\mathcal{P}(N)$ are uncorrelated (Baayen and Lieber, 1991).

Shen and Baayen (2021) applied these measures to two-syllable compounds in Mandarin Chinese. They selected compounds sharing a given constituent in a given position (initial or final), which they refer to as the ‘pivot’ constituent. For a range of pivots, Shen and Baayen (2021) observed a negative correlation between $V(N)$ (the number of compound types in their corpus that shared the same pivot) and the growth rate $\mathcal{P}(N)$ of the pivot’s morphological family. They also investigated the semantic transparency of pivotal morphological families. Using distributional semantics (Landauer and Dumais, 1997), they calculated how similar the meaning of a pivot is to the meaning of the compound in which this pivot occurs. The mean of this similarity, averaged over all compounds in the pivot’s morphological family, turns out to be positively correlated with the probability $\mathcal{P}(N)$ that with further sampling, previously unseen compounds with that pivot will be observed. Shen & Baayen (this volume) replicated these results for a larger set of pivots. They also report that a relatively recent powerful unsupervised clustering technique from machine learning, t-distributed stochastic neighbor embedding (t-SNE, Maaten and Hinton, 2008), does not detect clusters in semantic space for pivots, whereas for Mandarin suffixes, it does detect some clustering.

The present study reports research on German derivation that unfolded in part in parallel with the abovementioned research on compounding in Mandarin. Of central interest to us is the productivity of German particle verbs. Just as compounds consist of two (or more) words that can be used independently, particle verbs combine a particle and verb, both of which are independent words. A specific property of particle verbs is that in some constructions, the particle and the verb form one word, whereas in other constructions, other words can intervene between the two.

In what follows, we report that particle verbs have similar properties as the pivots of Mandarin compounds. Specifically, we replicate (1) the negative correlation between $V(N)$ and $\mathcal{P}(N)$, (2) the positive correlation between semantic transparency and $\mathcal{P}(N)$, and (3) the absence of by-particle clustering of particle verbs in the two-dimensional map constructed by t-SNE. Furthermore, for German suffixed words, we observed (1) the absence of a correlation between $V(N)$ and $\mathcal{P}(N)$, (2) a clear contribution of the suffix to the meaning of the derived word, and (3) well-formed clusters of derived word by suffix in the t-SNE map.

In order to better understand these differences between particles and suffixes, we made use of a combination of principal components analysis and linear discriminant analysis. It turns out that the information about the meanings of particles is spread out widely and thinly in distributional space, whereas the meanings of suffixes come with a strong signal that is well-represented already in the most highly ranked dimensions of a PCA orthogonalization.

This difference between particles and affixes was further supported by an investigation of the semantic relations between base and derived words. Using the CosClassAvg model (Shafaei-Bajestan et al., 2022, Shafaei et al., this volume), we constructed models that move from base word to derived word using vector addition to represent the semantics of particles and suffixes. As expected, the model for particles underperformed substantially compared to the model for affixes. As it is possible that vector addition is too restricted to model the semantic contribution of specifically particles with sufficient precision, we made use of a generalization of the FRACCS model (Marelli and Baroni, 2015), henceforth FRACCS++. This model also showed much better performance for affixes than for particles. Furthermore, it underperformed for both

¹In what follows, we use ‘types’ to denote distinct words, and ‘tokens’ to denote the occurrences of these words. Thus, in the sentence *And the cat sat on the mat*, there are 6 types and 7 tokens. The type *the* occurs twice — its token frequency is 2.

particles and suffixes compared to the CosClassAvg model, consistent with the results of Shafaei-Bajestan et al. (2022) and Shafaei et al. (this volume).

In the remainder of this study, we provide further detail on the abovementioned results. We first report our findings concerning the relation between semantic transparency and productivity. We then present our results obtained with t-SNE, after which we provide further details on the performance of the FRACCS++ and CosClassAvg models. Our study concludes with a discussion of the implications of our research.

2 Productivity

Many studies have proposed criteria for establishing whether a derivational rule is productive, and what factors co-determine its productivity (Schultink, 1961; Booij, 1977; Baayen, 2005; Plag, 1999; Dressler & Ladányi, 2000; Dressler, 2003; Fernández-Domínguez, 2009). What emerges from these studies is that, qualitatively, productivity decreases as the number of restrictions and conditions on a word formation process increases. Quantitative studies of productivity have led to several measures that provide an overall numeric assessment of different aspects of productivity (e.g., Baayen, 2005), but these measures by themselves cannot answer the question of to what extent the different qualitative factors contribute to the actual values of the quantitative measures.

The present study addresses one specific qualitative factor that has been argued to co-determine productivity, semantic transparency (see, e.g., Aronoff, 1976; Baayen, 1993; Bonami and Paperno, 2018). The difference between semantic transparency and semantic opacity is also referred to in the literature as the difference between non-lexicalized and lexicalized words (Lieber, 2010). For lexicalized words, the meaning of the complex word cannot be well predicted from the meanings of its constituents. Lieber gives as examples English *oddity* and *locality*. The former does not mean ‘the state of being odd’, but instead refers to a thing or person that is odd. Likewise, *locality* does not denote the state of being local, but a place or area. The greater productivity of English *-ness* as compared to *-ity* (Aronoff, 1976; Baayen and Renouf, 1996) is at least in part due to *-ness* having much more predictable semantics (see Riddle, 1985, for detailed discussion). A discussion of semantic transparency in the context of compounding can be found in Günther and Marelli (2019).

Our study explores ways in which the effect of semantic transparency on productivity can be measured quantitatively. Our study builds on a study of Mandarin adjective-noun compounding (Shen and Baayen, 2021) that reported that the category-conditioned degree of productivity \mathcal{P} (the hapax to token ratio $V(1, N)/N$, see Baayen (2001)) varies with several quantitative measures of semantic transparency: the tighter the meaning relation between an adjective and its compounds is, the greater the degree of its productivity is.

2.1 Extent of use and potential productivity

In this study, we consider two quantitative measures of morphological productivity. The first measure represents the number of different types $V(N)$ observed in a sample of N tokens. We interpret this measure as estimating the ‘extent of use’ or the ‘profitability’ of a morphological category (cf. Baayen, 2005; Corbin, 1987). The second measure, which is a straightforward application of the Good-Turing estimate of the probability of unseen types (Good, 1953), assesses the likelihood that if the sample is extended, novel types will be found. These two measures can be related to each other through the urn model (or bag of words model) for word frequency distributions. If $V(N)$ denotes the number of types observed for N tokens, then the tangent to $V(N)$ (conceptualized as a function of N) is given by the (expectation of) the ratio of hapax legomena $V(1, N)$ and the number of tokens N :

$$\mathcal{P}(N) = V(1, N)/N.$$

When increasingly large samples are taken from the same population, then $V(N)$ is a monotonically increasing, but decelerating, function of N . Accordingly, the growth rate $\mathcal{P}(N)$ decreases as N (and $V(N)$)

increase. For completely unproductive morphological categories, $V(N)$ will be constant from a certain sample size onwards, and correspondingly, $\mathcal{P}(N)$ will be zero. (For mathematical details, the reader is referred to Baayen, 2001).

If morphological categories (e.g., words with the particle *weg*, ‘away’, or the prefix *un-*, ‘un-’) are sampled from the same underlying distribution, then one would expect that $V(N)$ and $\mathcal{P}(N)$ enter into a negative correlation. In case that they are sampled from different populations, no correlation can be predicted.

Let us now consider German particle verbs and compare these with German affixed words. German particle verbs (e.g., *backen* ‘bake’, *durchbacken*, ‘bake through’, *kurz*, ‘short’, *abkürzen*, ‘shorten’) distinguish themselves from standard prefixed verbs by the property that the particle can occur separated from its base verb. This happens in main clauses without auxiliary verbs, in which case the particle can follow the base verb immediately, but also can be separated from the base verb by one or more words. The past participle of particle verbs inflects with *ge-* appearing between particle and verb. In infinitive clauses, *zu* appears in between the particle and its base verb. Whereas some particle verbs are fairly transparent, such as *weglaufen* (away walk, ‘to walk away’), others are not so straightforward to understand on the basis of their constituents, for instance *abtragen*, (away-carry, which has meanings as diverse as ‘pay off’, ‘ablate’, and ‘remove’). Because particle verbs have constituents that both can be used as independent words, they can be considered to be compounds rather than derived words.

The upper panel of Figure 1 plots, on a log-log scale, potential productivity $\mathcal{P}(N) = V(1, N)/N$ against the number of types $V(N)$ (realized productivity, or ‘extent of use’), for a total of 98 particles. Type and token frequencies were taken from dlexDB (<http://dlexdb.de/>). Clearly, particles that have given rise to more types (larger $V(N)$) tend to be less productive in terms of $\mathcal{P}(N)$. A similar negative correlation was also observed by Shen and Baayen (2021) for Mandarin adjective-noun compounds. Apparently, an increase in realized productivity is detrimental for the potential productivity of German particle verbs.

Shen and Baayen (2021) pointed out that English derivational suffixes are characterized by the absence of a correlation of $V(N)$ and $\mathcal{P}(N)$ (see also Baayen and Lieber, 1991). The lower panel of Figure 1 clarifies that the same holds for German affixal derivation. A set of 18 German derivational affixes (of which 4 are prefixes, highlighted in red; frequencies extracted from the CELEX database), are scattered unsystematically in the $V(N) - \mathcal{P}(N)$ plane.

One possible explanation for the pattern that emerges from Figure 1 is that the particle verbs are sampled from roughly the same general population, whereas the derived words come from separate, affix-specific, populations. In the remainder of this study, we address this possibility by investigating the transparency of these word formation processes using distributional semantics. We first investigate measures of transparency that compare constituent semantic vectors with the semantic vectors of the complex words.

In the present study, we made use of pre-trained word embeddings, obtained with the fasttext algorithm (Bojanowski et al., 2017), that are available at <https://dl.fbaipublicfiles.com/fasttext/vectors-crawl/cc.de.300.vec.gz>.

2.2 Semantic transparency assessed with constituent vectors

Semantic transparency is one of the many factors co-determining productivity (see, e.g., Aronoff, 1976; Baayen, 1993; Bonami and Paperno, 2018). One way of assessing the semantic transparency of a complex word is to compare the semantic vector of one of the constituents of a complex word with the semantic vector of the complex word itself. For particle verbs, we have two options. We can compare the semantic vector of the particle with the semantic vector of the particle verb. Henceforth, we refer to this measure of transparency as T_p . Alternatively, we can compare the semantic vector of the base word with the semantic vector of the particle verb, henceforth T_b . For affixed complex words, we can only compare the semantic vector of the base word with that of the complex words as, by definition, affixes are bound forms for which we do not have independent word embeddings.²

²Fasttext can be used to obtain semantic embeddings for sub-word n-grams (for n ranging from 3 to 6). There are three reasons for not using these subword embeddings. First, our general approach is set up to be compatible with other embeddings such as provided by GLOVE, word2vec, and the grounded embeddings developed by Shahmohammadi et al. (2021). Second, an evaluation of the substring embeddings of fasttext for morphology is far from trivial. For instance, should we take *#be* as

We first consider T_p , the semantic similarity of the particles and the corresponding particle verbs. For each pair of particle and particle verb, we calculated the Pearson correlation of their semantic vectors. (An alternative measure that is commonly used is the cosine of the angle of the two vectors. As the cosine measure and the correlation measure are mathematically related and tend themselves to be almost perfectly correlated, the choice of measure does not affect results. We opted for the correlation measure, due to its statistical properties. For instance, with the correlation measure, we can evaluate whether two vectors are significantly correlated.) Subsequently, we grouped the resulting correlations by particle, and calculated the by-particle mean correlations. Figure 2 plots $\log \mathcal{P}$ against $\log T_p$ (the mean correlation of the semantic vectors of the particle and the particle verb). We observe a positive correlation between the two measures, similar to the positive correlation of the semantic vectors of the adjectives and the vectors of the corresponding adjective-noun compounds in Mandarin (see Shen and Baayen (2021) and Shen & Baayen, this volume).

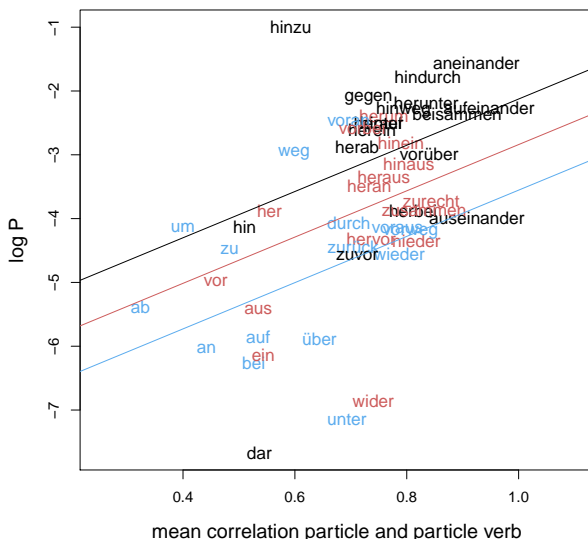


Figure 2: Scatterplot for potential productivity \mathcal{P} and T_p , the by-particle means of the correlations of the embeddings of particles and the embeddings of the corresponding particle verbs. . Color coding distinguishes between particles with one meaning (black), two meanings (red) and three meanings (blue). Degree of productivity decreases with number of meanings ($\hat{\beta} = -0.714, p = 0.0019$), and increases with the mean correlation of particle and particle verb ($\rho = 0.23, S = 1641802055, p < 0.0001$).

In the scatterplot, shorter particles (e.g., *ab*) dominate in the lower left, and longer particles (e.g., *aneinander*) dominate in the upper right. However, particle length (in letters) turned out not to be predictive for \mathcal{P} . A predictor that did receive good support is the number of meanings that are listed for the particles in the DWDS (Digitales Wörterbuch der deutschen Sprache, Berlin-Brandenburgischen Akademie der Wissenschaften, <https://www.dwds.de/d/wb-dwdswb>). Particles with three meanings (blue line) tend to have lower values of T_p , as expected, and also are of lower productivity. Conversely, particles with one

the proper embedding for the prefix *be-*, or should we take a (possibly) weighted average of *#be* and all trigrams *beX*? It is also unclear how to compare embeddings for short and long affixes, as the longer ones can be built up from many more partial n-grams. Third, fasttext substring embeddings are likely to be extremely noisy. Schreuder and Baayen (1994) observed for prefixes that words starting with a pseudo-prefix (e.g., *re* in *reindeer*, or *un* in *uncle*) are so frequent that their cumulative token frequency is larger than the cumulative token frequency of the words that actually are prefixed. Since fasttext is token-based, the substring vectors for prefixes will be substantially contaminated by the many tokens of pseudo-prefixed words. A more principled way of obtaining embeddings for affixes is presented in Baayen et al. (2019), but this study implemented affixal vectors only for a small corpus.

meaning only (black line) have higher values of T_p , and they also have higher values of \mathcal{P} . The regression lines for particles with one, two, and three meanings are based on a Gaussian Location-Scale (gaulss) linear model. We used a gaulss model because the scatterplot suggests that the variance of the response is not homoskedastic in the predictor. A gaulss model fitted with the `gam` function of the `mgcv` package revealed for the mean that \mathcal{P} decreased with the number of meanings ($\hat{\beta} = -0.71, p = 0.0019$) and that \mathcal{P} increased with T_p ($\hat{\beta} = 3.62, p = 0.0169$). The variance, by contrast, decreased with T_p ($\hat{\beta} = -1.78, p = 0.0374$).

With respect to the second transparency measure, T_b , we did not observe any evidence for a correlation with $\mathcal{P}(N)$, as shown in the upper left panel of Figure 3. The lower left panel of this figure clarifies that a correlation is actually present for T_b and the number of types $V(N)$ ($r = 0.64, t(38) = 5.17, p < 0.0001$). The right-hand panels of Figure 3 illustrate that affixed words are characterized by the same quantitative patterns: no correlation of the transparency measure T_b with $\mathcal{P}(N)$, but a decent correlation of similar magnitude ($r = 0.65, t(16) = 3.39, p = 0.0037$) is present for T_b and $V(N)$. Apparently, the more semantically similar the base words and the derived words that belong to a given morphological category are, the greater their extent of use is, and the more profitable (Corbin, 1987) that category is. However, this measure of semantic transparency is, surprisingly, not predictive for potential productivity $\mathcal{P}(N)$.³

3 Geometry of semantic transparency in semantic space

The two measures of transparency that we considered thus far, T_p and T_b , focus on the semantic relations of individual pairs of constituents and their corresponding carrier words. In what follows, we address the geometry of semantic transparency in semantic space.

3.1 t-SNE analysis of semantic space

We first explore whether, and if so, how, the semantic vectors of complex words cluster in semantic space. Next, we investigate how the shift vectors (the vectors that start at the vector of the base and end at the vector of the complex word) cluster in shift space.

To do so, we made use of t-distributed stochastic neighbor embedding (Maaten and Hinton, 2008). t-SNE is a nonlinear dimension reduction technique that allows us to represent the semantic vectors of complex words in a two-dimensional plane, such that, with high probability, words that have similar semantic vectors will be close together, and words that have dissimilar semantic vectors will be far apart. If there are clusters in a high-dimensional space, t-SNE is likely to find these. For our cluster analyses, we have used the default settings of the `Rtsne` package (Krijthe, 2015) for R. The output of a t-SNE analysis is a table with the coordinates of the observations in a reduced two (or three) dimensional space. Since we know the morphological structure of the observed words, we use visualization with color highlighting to clarify whether the t-SNE indeed succeeded in placing words with similar morphological structure in each other’s vicinity.

Figure 4 presents the results of a t-SNE cluster analysis of those 22 particles for which we have semantic vectors for at least 10 particle verbs. The reason for setting a threshold on the minimum number of verbs is that the interest of this analysis is in potential clusters. For very small numbers of types, clusters are unlikely to be meaningfully traceable. Figure 4 clarifies that only very few clusters are detected, namely, for *zurück* (light brown, upper left), *zusammen* (grey, to the lower left of the origin), *herum* (pink, far left), and *durch* (dark pink, slightly above *zusammen*).

Figure 5 presents the results of a t-SNE cluster analysis for the derivational affixes. All derivational suffixes show considerable clustering. Of the prefixes, only *un-* shows clustering. By contrast, the prefixes *be-*, *ver-*, and *ent-* are found in one large cluster in the right-hand side of the scatterplot. The subcluster of brown datapoints in the upper right of this cluster represents words with the suffix *-isier*.

In summary, most particle verbs, as well as the prefixes *be-*, *ver-*, *ent-* do not cluster in semantic space. By contrast, derivational suffixes and the prefix *un-* appear with reasonably well-defined clusters. Surprisingly, a linear discriminant analysis (LDA) can assign semantic vectors to both affixes and particles with a high

³In order to back of from zero, $\mathcal{P}(N) = V(1, N)/N$ was transformed into $(V(1, N) + 1)/(N + 1)$ before taking the logarithm.

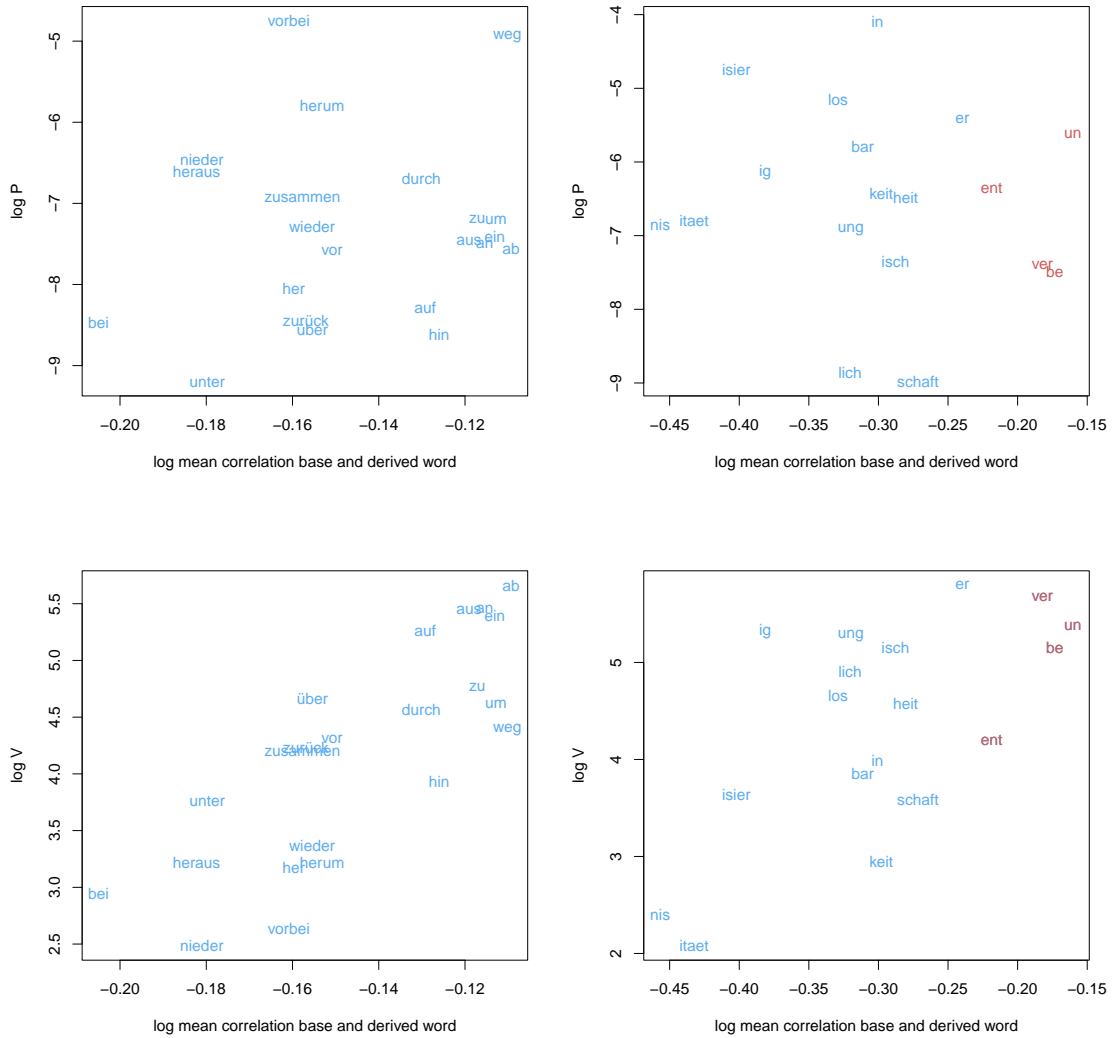


Figure 3: Log mean correlation of the semantic vectors of base word and complex word (horizontal axis) for particles verbs (left panels) and affixed words (right panels), by log potential productivity ($\mathcal{P}(N)$, upper panels) and extent of use ($V(N)$, lower panels). Prefixes are highlighted in red.

degree of accuracy (94%) when the model is given access to all 300 dimensions of the semantic vectors. This indicates that there must be some structure in semantic space also for the particle verbs.

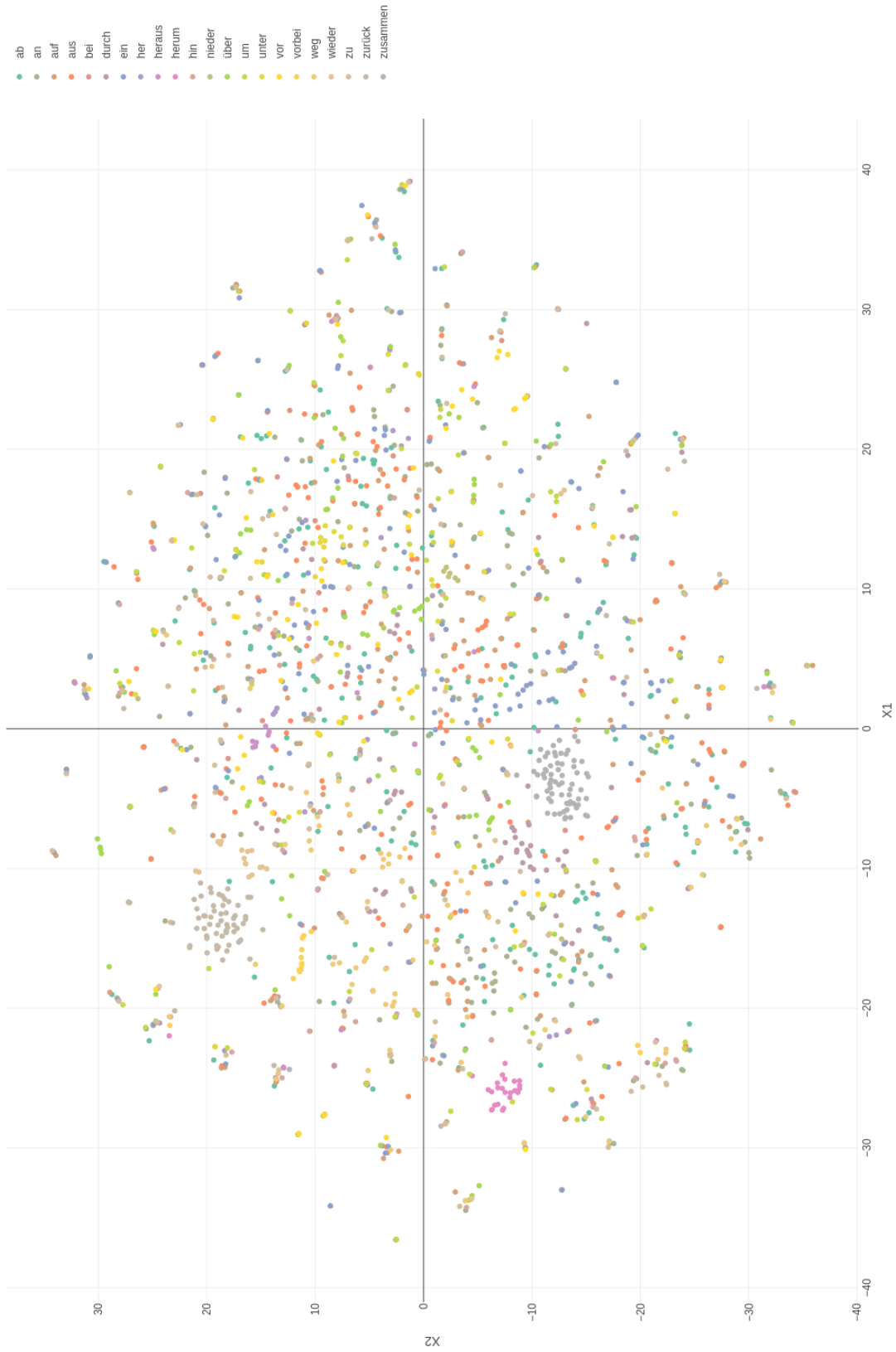


Figure 4: Locations in t-SNE space of words with the 22 particles, each of which is represented by at least 10 verbs. Only a few clusters are visible: *zurück* (light brown, upper center), *zusammen* (grey, to the lower left of the origin), *herum* (pink, far left), and *durch* (dark pink, slightly above *zusammen* ([interactive plot here](#))).

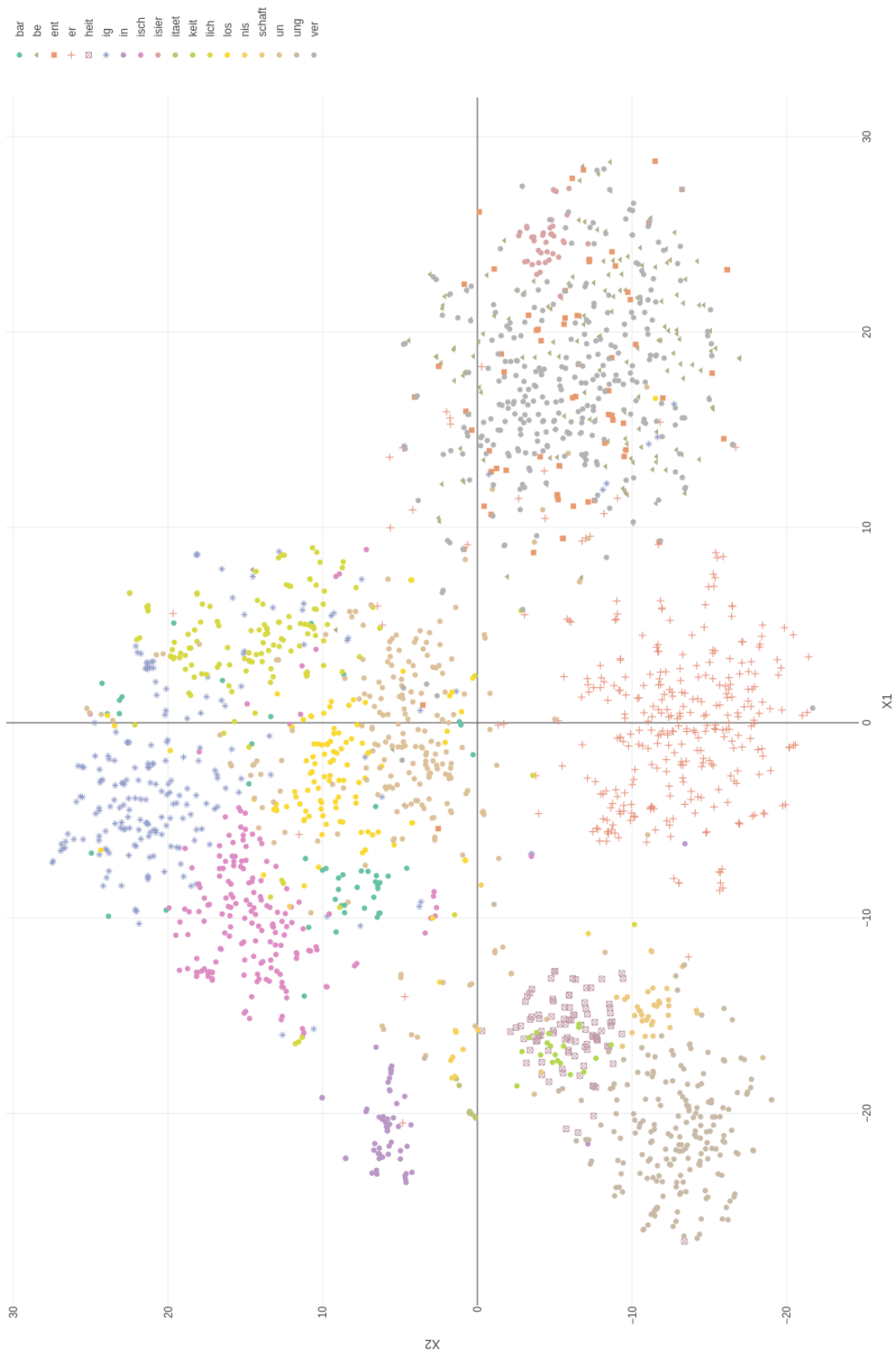


Figure 5: Locations of words with 18 derivational suffixes in t-SNE space. Derivational suffixes show considerable clustering, with as exceptions the prefixes *be-*, *ver-*, and *ent-*, which are found in the same large cluster in the upper central part of the scatterplot. The embedded pink cluster represents words with the suffix *-isier* ([interactive plot here](#)).

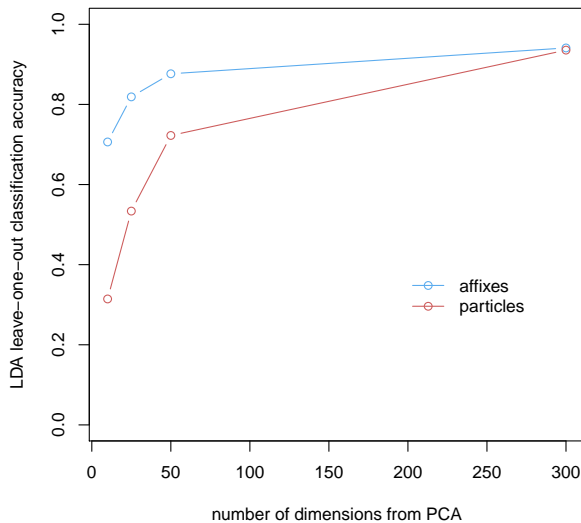


Figure 6: Leave-one-out cross-validation accuracy of linear discriminant analyses applied to the first $n = 10, 25, 50, 300$ dimensions of PCA-orthogonalized semantic vectors for words with affixes (blue) and particles (red). For particles, accurate classification requires more principal components.

In order to understand why the t-SNE analysis does not detect this structure for the particle verbs (and three verbal prefixes), we carried out a principal component orthogonalization of the semantic space. The dimensions of this new space are ordered such that the first dimension captures most of the variance in the original fasttext space, and the last dimension the least of this variance. We then carried out a series of LDA analysis that were given access to the first 10, the first 25, and the first 50 dimensions, as well as to all 300 dimensions.

As shown in Figure 6, while classification accuracy is around 94% for both affixes and prefixes when the full 300-dimensional space is used, classification accuracy decreases quickly when only the first n most important principal components are available to the classifier. Importantly, classification accuracy decreases more quickly for words with particles (red) than for words with affixes (blue) as more dimensions are withheld. For instance, when only the top 10 principal components are made available to the LDA, accuracy is at 71% for the affixes but only at 31% for the particle verbs. (The cumulative proportions of variance explained by the two PCA orthogonalizations are, for the affixes, 0.36, 0.50, 0.63, and 1, and for the particles, 0.30, 0.46, 0.61, and 1. Important for the above comparisons is that the cumulative proportions are of similar magnitude. It is noteworthy, however, that the proportions for the particles lag slightly behind those for the affixes, consistent with the affixes having more distinct distributions in semantic space.)

From the fact that the semantic vectors of affixed words are much more strongly represented on the first 50 principal components than is the case for particle verbs, it follows that the semantic vectors of affixed words capture more variance and are more informative than the vectors of particle verbs.

As always in scientific inquiry, absence of evidence does not constitute evidence of absence. The absence of clusters for particle verbs does not imply their semantics are completely random. There must be information in the embeddings of particle verbs: a supervised classifier can predict with high accuracy what particle is present, provided that the classifier has access to all 300 dimensions. However, compared to suffixed words, this information is hidden from view for unsupervised learning. Suffixed words have distributional properties that are visible already in the top principal components of a PCA orthogonalization, indicating that their similarities and dissimilarities play a much more pervasive role in structuring the semantic space.

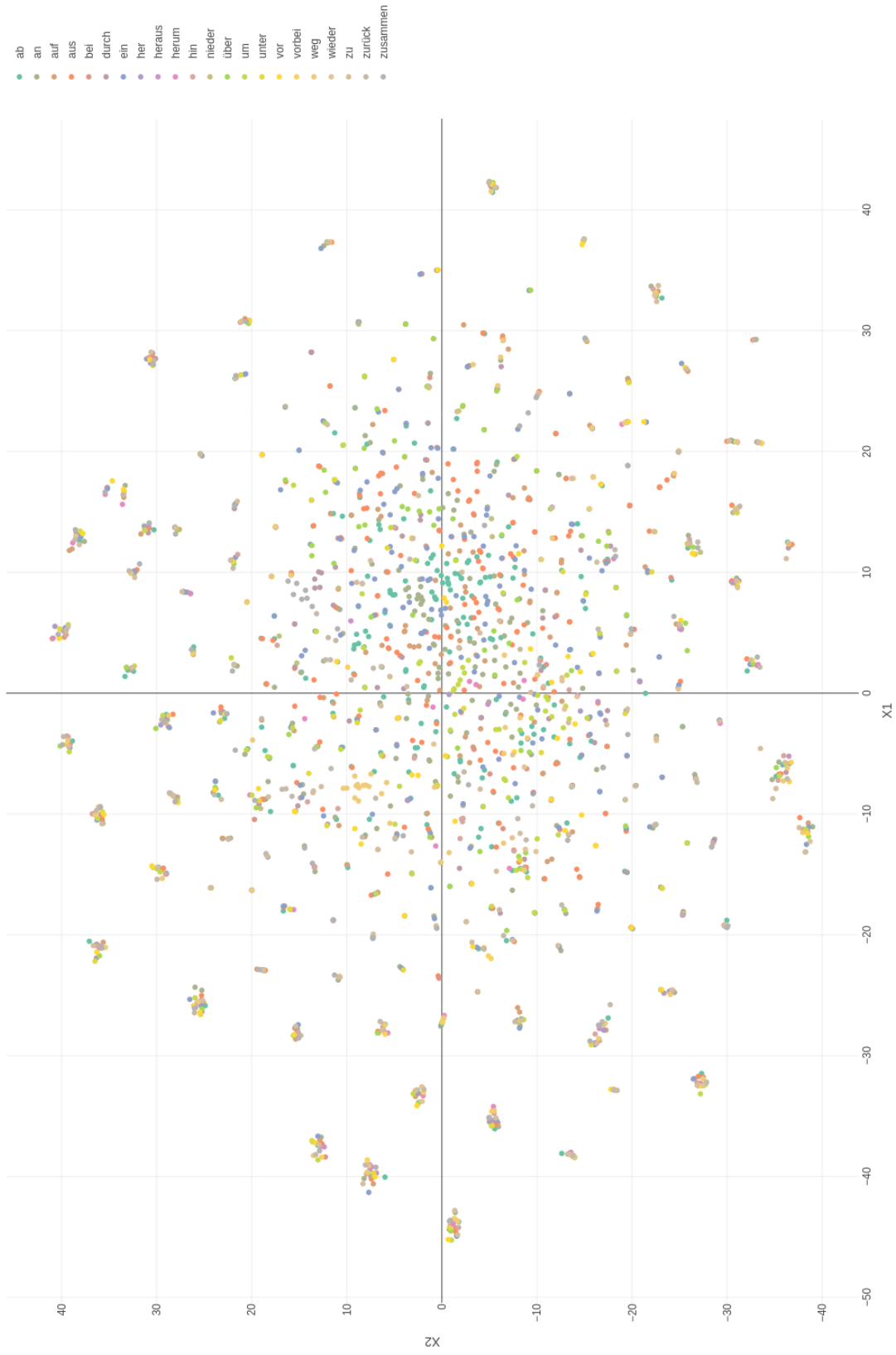


Figure 7: Locations of the shift vectors for particle verbs in t-SNE space. The clusters in the periphery of the scatterplot pertain to particle verbs sharing the same verb stem ([interactive plot here](#)).

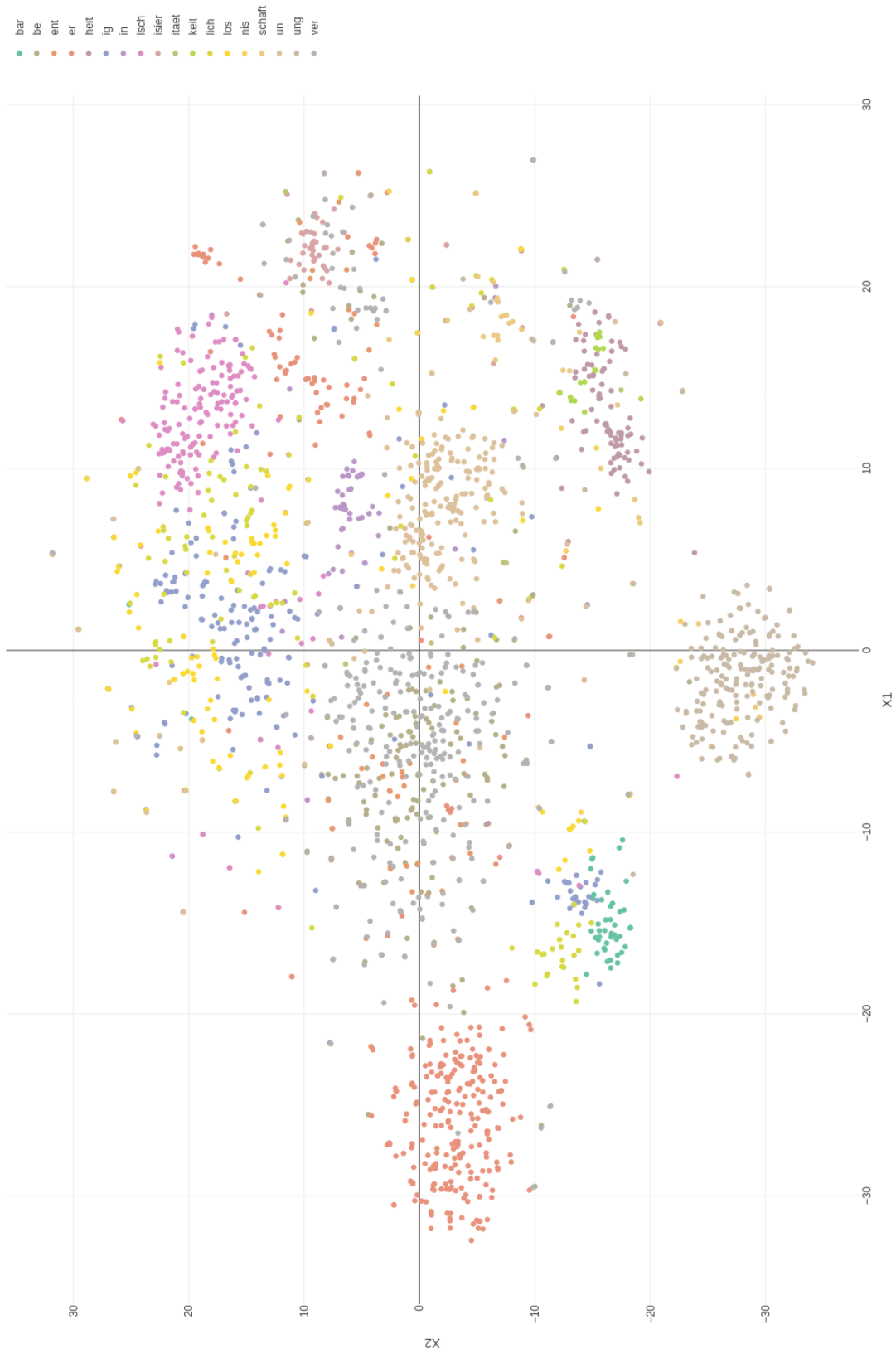


Figure 8: Locations of the shift vectors of words with derivational suffixes in t-SNE space. Data points cluster reasonably well by suffix ([interactive plot here](#)).

3.2 t-SNE analysis of shift vectors

We have seen that the vectors of particle verbs are scattered in the two-dimensional characterization of semantic space generated by the t-SNE algorithm. It is conceivable that this scatter is a property of their base words, and that a given particle implements the same, or a similar, semantic change for all base words. To examine this possibility in more detail, we first examined the shift vectors of particle verbs. Below, we will consider more general, less restricted mappings between base and derived words.

Shift vectors are those vectors which when added to the vectors of the base verb, result in the vector of the particle verb:

$$\overrightarrow{\text{AUFHÖREN}} = \overrightarrow{\text{HÖREN}} + \underbrace{(\overrightarrow{\text{AUFHÖREN}} - \overrightarrow{\text{HÖREN}})}_{\text{shift vector}}.$$

If adding a particle or affix to a base word implements the same translation in semantic space, then the shift vectors for all complex words with the same particle or affix should be the same. Given that embeddings are likely to contain a noise component, the shift vectors for a given formative are likely to show some variability around a centroid. However, if the shift vectors are dominated by semantic idiosyncrasies of the complex words, then the shift vectors are expected to show a wide scatter in shift space, without forming clear clusters.

Figure 7 presents a t-SNE map for the shift vectors of the particle verbs. The color coding is by particle. Clearly, the shift vectors for a given particle show a wide scatter. There are some small clusters, mainly in the periphery of the t-SNE space, which turn out to represent complex words that share the same base words. Occasionally, a cluster contains particle verbs derived from near-synonymous base words (e.g., *kriegen*, *bekommen*). This clarifies that the geometric relation between base and derived word cannot be captured in an insightful way with shift vectors.

Do the shift vectors of affixes cluster in t-SNE space? Figure 8 shows that the shift vectors for German suffixes indeed form relatively good clusters. There is some overlap as well as some scatter for individual complex words. However, by and large, the t-SNE analysis suggests that suffixes have two properties. First, suffixes contribute their own specific shift vectors, which we can represent as the vectors pointing to the centroids of their respective clusters. Second, variances in the angle around the centroid appear to be pretty much the same, and appear not to co-vary with the number of types V ($p > 0.2$). This is consistent with a model in which the semantic vectors of complex words have random error terms with the same variance (cf. Nikolaev et al., this issue).

Figure 9 presents the shift vectors for prefixed words in the same t-SNE plane as that of Figure 8. The only prefix that shows good clustering is *un-*, which is perhaps unsurprising given its straightforward semantic contribution of negation (e.g., *klar/unklar*, ‘clear/unclear’). The prefixes *be-*, *ver-*, and *ent-* are characterized by a wide scatter and enormous overlap. Although at a high level of abstraction, these prefixes can be differentiated semantically (see, e.g., Lieber and Baayen, 1993, for the very similar Dutch prefixes *be-*, *ver-* and *ont-*), Figure 9 clarifies that individual complex words have strongly idiosyncratic semantics and that current methods from distributional semantics are not sensitive, possibly cannot be sensitive, to the high-level abstract meanings of these prefixes.

In summary, the relation between base and derived words can be captured by shift vectors for suffixes and the most transparent prefix (*un-*), whereas particles and the prefixes *be-*, *ver-* and *ent-* do not show consistent clustering. In the next section, we address the question of whether a more complex mapping from base to derived word yields improved results, not only for these three prefixes, but also for particles.

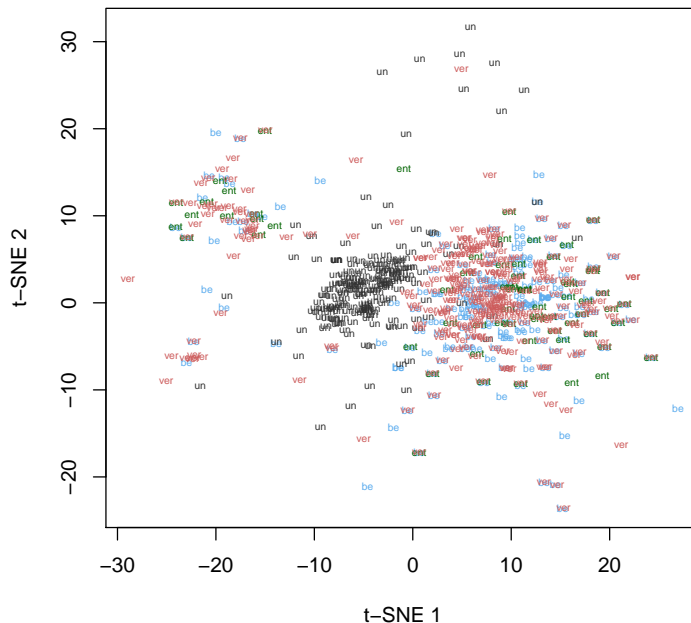


Figure 9: Locations of the shift vectors for prefixed verbs in t-SNE shift space. The only prefix revealing a dense cluster (but in combination with a wide scatter outside this cluster) is *un-*.

4 Modeling conceptualization: FRACSS++ and CosClassAvg

We have seen that most particles, as well as the prefixes *be-*, *ver-* and *ent-*, do not show any clustering in semantic space. We have also shown that a linear discriminant analysis can predict particles or affixes with a high accuracy, also under cross-validation, but that having access to all 300 semantic dimensions is crucial. The final question we address in this study is how well the individual vectors of complex words can be predicted from the vectors of their base words. The LDA analysis was given the task to assign a vector to a category label. In what follows, we consider a more difficult task, namely, to predict, as accurately as possible, the 300-dimensional vector of the complex word, given the vector of its base word. In other words, we are asking whether we can quantify the conceptualization of the meanings of particle verbs and affixed verbs.

Our point of departure is the FRACCS model of [Marelli and Baroni \(2015\)](#). The FRACCS model was originally developed for individual affixes. To implement this model, we take as starting point a matrix \mathbf{B} with the semantic vectors of all the words with a given affix, and we estimate a linear transformation \mathbf{F} that maps these base vectors onto the semantic vectors of the corresponding complex words, brought together as the row vectors of a matrix \mathbf{D} :

$$\mathbf{D} = \mathbf{B}\mathbf{F}.$$

Using the Moore-Penrose pseudo-inverse of \mathbf{B} , denoted by \mathbf{B}^+ , \mathbf{F} is obtained straightforwardly:

$$\mathbf{F} = \mathbf{B}^+\mathbf{D}.$$

Given \mathbf{F} , we obtain the predicted semantic vectors

$$\hat{\mathbf{D}} = \mathbf{B}\mathbf{F};$$

if the mapping is precise, a predicted vector $\hat{\mathbf{d}}_i$ should have as its closest neighbor the target gold-standard vector \mathbf{d}_i .

A fundamental problem with this model is that although it will learn the words on which it is trained very well, it cannot generalize. For instance, there are 173 words with *be-* in our dataset. When we set 156 words apart for training, and a random 17 words for testing, the FRACCS model is 100% accurate for the training data, but only gets one out of 17 held-out words correct (6% accuracy). Clearly, FRACCS works well as a memory, but it is a memory that is not productive. The reason for this overfitting is straightforward: there are only 173 observations, each represented by 300 real numbers. The FRACCS mapping matrix is of dimension 300×300 , which amounts to 90,000 parameters. Thus, from a statistical and machine learning perspective, the number of parameters and the number of observations are completely out of balance.

In order to attenuate this problem somewhat by increasing the number of observations, we applied the FRACCS method to all the 2203 affixed words in our study jointly, and likewise to the set of 2019 particle verbs. This resembles the approach to compounding reported by [Marelli et al. \(2017\)](#), where one model mapping is applied to a heterogeneous set of compounds that exemplify a wide range of semantic relations between the modifier and head constituents. Of course, the FRACCS method is now confronted with a more complex task, and accuracy is now unlikely to remain at 100%. As from a research perspective, having a series of overfitting models that all perform with close to 100% accuracy without generalizing is not informative at all, our extension of FRACCS, to which we will refer as FRACCS++, is a more useful tool with which we can probe the relations between the meanings of the base words and the corresponding derived words.⁴

We first applied FRACCS++ to the 2203 affixed words, and observed that for 62.4% of these derived words, the closest neighbor of the predicted vector was indeed the targeted semantic vector. The left half of [Table 1](#) provides examples of the kind of prediction errors made by this mapping. When the words with the prefixes *be-*, *ver-* and *ent-* are removed from the dataset before estimating \mathbf{F} , accuracy increases from 62.4% to 75.9%. This result dovetails well with the observation that these three prefixes do not show any clear clustering in the t-SNE plane.

| affixed verbs | | particle verbs | |
|------------------|----------------|----------------|-----------------|
| target | predicted | target | predicted |
| abenteuerlich | abenteurer | angehören | zusammengehören |
| achter | achtlos | anfertigen | abfertigen |
| adelig | adlig | anknüpfen | aufknüpfen |
| akrobatin | akrobatisch | anmerken | vormerken |
| altern | veralten | anstrengen | überanstrengen |
| araber | arabisch | andauern | überdauern |
| artig | unart | anklagen | einklagen |
| aerztin | verarzten | ankämpfen | durchkämpfen |
| atmung | beatmen | anfechten | ausfechten |
| ausbeuter | ausbeutung | anstarren | hinstarren |
| ausgeglichenheit | unausgeglichen | anhäufen | überhäufen |
| automatisch | automatisieren | anfeuern | abfeuern |
| banner | verbannen | anraten | abraten |
| beachten | verachten | anfühlen | einfühlen |
| beantworten | verantworten | anreizen | ausreizen |

Table 1: Examples of words incorrectly predicted by FRACSS++ for affixed words (left) and particle verbs (right)

For particle verbs, we expect accuracy to be substantially reduced, as most particle verbs also do not

⁴We believe that FRACCS++ is a straightforward extension of the idea of FRACCS, but as one of our reviewers insisted that FRACCS is a model that is defined for individual affixes only, we refer to our generalization with a different name that, however, reveals our indebtedness to the FRACCS model.

show any clear clusters. Accuracy for a mapping trained on the 2220 base verbs of particle verbs was at 36.4. The right half of Table 1 illustrates the kind of errors produced by this mapping. Considered jointly, these results indicate that a wide scatter and the absence of clustering in the t-SNE plane on the one hand, and a low accuracy when predicting complex words from base words using FRACSS++ on the other hand, go hand in hand.

Next, we adopt the approach of Shafaei-Bajestan (Shafaei-Bajestan et al., 2022), and use the CosClassAvg method to model the conceptualization of complex words. Above, we presented t-SNE maps for the shift vectors of particles (Figure 7) and of affixes (Figure 8). The latter figure shows clustering of shift vectors by affix, with as exceptions the prefixes *be-*, *ver-*, and *ent-*. These clusters suggest that the affixes tend to change the meaning of a base word into the meaning of the corresponding derived word using a consistent shift, i.e., by adding an affix-specific vector. The CosClassAvg formalizes this intuition by predicting the meaning $\hat{\mathbf{d}}_i$ of word i from the meaning of its base word \mathbf{b}_i to which an affix-specific (or particle-specific) shift vector \mathbf{m}_j is added. The shift vector of a given affix (or particle) is obtained by averaging over all the individual shift vectors of the words sharing affix or particle m_j :

$$\hat{\mathbf{d}}_i = \mathbf{b}_i + \mathbf{m}_j.$$

For the affixed words, accuracy was 72.2% when the prefixes *be-*, *ver-* and *ent-* are included. Accuracy increased to 78.8% when these three prefixes are excluded. For the particle verbs, accuracy was at 50.3%.

In summary, the CosClassAvg method outperforms FRACSS++ with respect to accuracy for both affixed words (72.2% vs. 62.4%) and for particle verbs (50.3% vs. 36.4%). Furthermore, CosClassAvg has somewhat improved performance for unseen data. For training on 90% of the data, and predicting for 10% of the data, accuracy for CosClassAvg for the affixed words was 65% for training and 49% for testing (a reduction of accuracy for the held out data to 75%). For the FRACSS++ model, the corresponding accuracies were 55% and 28% (a reduction to 51%).

These results raise the question of why CosClassAvg performs better for the present data than the FRACSS++ baseline model, not only for the affixed words, but also for the particle verbs. To answer this question, consider that CosClassAvg builds on the centroids of individual clusters. Conversely, the predictions of FRACSS++ are based on all words jointly. Thus, CosClassAvg is set up so that it is eminently sensitive to the semantics of a given morphological category, without interference from other morphological categories. This sensitivity to local structure in semantic space may also help explain why CosClassAvg has greater accuracy for particle verbs. Although particle verbs do not form clusters that are visible to the t-SNE algorithm, we know from the LDA analysis that in the full high-dimensional space, there is some structure. By conditioning on individual particles, apparently, irrelevant sectors of semantic space are not taken into consideration, resulting in somewhat better prediction accuracy. However, compared to the affixed words, prediction accuracy remains substantially reduced, indicating that particle verbs are generally less transparent than affixed words.

As a point measure of the clustering properties of affixed words and particle verbs, we calculated, for each affix and each particle separately, the mean of the pairwise correlations of the semantic vectors of all words sharing that affix or particle. We then compared this final transparency measure (which follows Shen and Baayen, 2021) with the two measures of productivity, $\mathcal{P}(N)$ and $V(N)$. Correlations with $\mathcal{P}(N)$ turned out not to be well supported, not for affixes nor for particles ($\rho = -0.295, p = 0.064$ for particle verbs, $\rho = -0.270, p = 0.2791$ for affixed words). This is perhaps unsurprising, as in these analyses, affixes and particles with less than three types are not taken into consideration. However, as illustrated in Figure 10, for particles (right panel), a strong negative correlation emerged ($\rho = -0.739, p < 0.0001$). By contrast, for affixes, there was no evidence for a correlation at all ($\rho = -0.125, p = 0.6208$). As can be seen in the right panel of Figure 10, longer particles are found more to the right. We therefore carried out a linear regression with number of letters of the particle as covariate. Only the transparency measure was supported, with no support for an effect of length ($p > 0.5$). This allows us to conclude that, apparently, a greater mean pairwise correlation, henceforth T_m predicts a lower extent of use V . In other words, the usefulness of a particle within a language community appears to be that it can be employed to create a wide range of rather different meanings. The more the meanings of a particle’s carrier words are similar to each other, the less

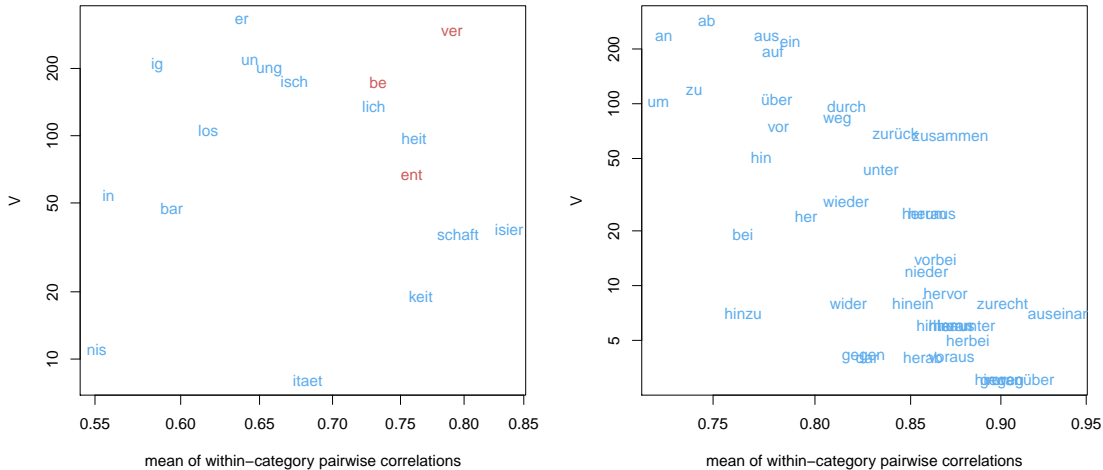


Figure 10: Mean of within-category pairwise correlations as predictor of the number of types. Left: affixed words; right: particle verbs.

likely it is that it will be used to coin more words.

5 Discussion

In this study, we investigated the relation between morphological productivity and semantic transparency for German particle verbs and derivational affixation. We considered two aspects of morphological productivity, profitability or extent of use, assessed with the type count $V(N)$, and potential productivity $\mathcal{P}(N)$, the Good-Turing estimate of sampling new types (Baayen, 2005). Following Shen and Baayen (2021), we considered three measures of semantic transparency based on distributional semantics. T_p is the mean of the correlations of the semantic vectors of particle and particle verb. T_b is the mean of the correlations of the semantic vectors of the base word and the complex word. Finally, T_m is the mean of all pairwise correlations of the semantic vectors of the complex words sharing a given particle or affix. We also investigated semantic transparency using exploratory visualization with t-SNE (Maaten and Hinton, 2008) as unsupervised clustering method. Finally, we compared two models for the conceptualization of the meanings of complex words: FRACSS++, building on the FRACCS model by Marelli and Baroni (2015), and CosClassAvg (Shafaei-Bajestan et al., 2022). The results that we obtained are summarized in Table 2.

Particle verbs, but not affixed words, revealed a negative correlation between $V(N)$ and $\mathcal{P}(N)$. This negative correlation is expected for data that are well-approximated by the urn model (or bag-of-words model) for word frequency distributions (Baayen, 2001): as more types are sampled, the rate at which new types can be sampled decreases. The presence of a negative correlation of $V(N)$ and $\mathcal{P}(N)$ for particle verbs suggests that the particles do not make highly distinct and well-differentiated contributions to the meanings of their carrier verbs. This finding dovetails well with the absence of by-particle clustering in semantic space, as gauged with t-SNE.

This result does not imply that the relation between the semantics of the constituents of particle verbs, and the semantics of the particle verbs themselves, is completely random. A linear discriminant analysis is able to predict the identity of the particle from the semantic vectors of the particle verbs with 94% accuracy. However, given the more difficult task to predict the semantic vector of a particle verb itself from the semantic vector of its base word, the FRACSS model reaches an accuracy of only 36%, and the CosClassAvg model an accuracy of 50%.

Table 2: Summary of the quantitative similarities and differences between particle verbs and affixed words.

| | particle verbs | affixed words |
|--------------------------------------|-----------------------|----------------------|
| V and \mathcal{P} | negative correlation | no correlation |
| T_p and \mathcal{P} | positive correlation | - |
| T_b and \mathcal{P} | no correlation | no correlation |
| T_b and V | positive correlation | positive correlation |
| T_m and \mathcal{P} | no correlation | no correlation |
| T_m and V | negative correlation | no correlation |
| LDA accuracy given 300 dimensions | 94% | 94% |
| LDA accuracy given 10 PCA dimensions | 31% | 71% |
| t-SNE words | no clustering | clustering |
| t-SNE shift vectors | some clusters by base | clusters by affix |
| FRACSS accuracy | 36.4% | 62.4% |
| CosClassAvg accuracy | 50.3% | 72.2% |

The T_p and T_b measures gauge the transparency of constituents with respect to their carrier words. The T_p measure, which is available only for the particle verbs, is positively correlated with \mathcal{P} : the more transparent the particle, the more likely it is that unseen particle verbs are present in the population. The T_b measure turns out to be positively correlated with V , suggesting that base transparency is essential for profitability. Importantly, the T_m measure, which evaluates all pairwise whole-word similarities, is negatively correlated with V . The more similar a particle’s verbs are, the less profitable a particle is.

Considered jointly, these results suggest that the productivity of a particle hinges on the ‘local’ transparency of constituents and particle verb, rather than on ‘global’ semantic changes that are realized across all words sharing a particle: The particle itself does not appear to make a systematic semantic contribution that generalizes across all the particle verbs it gives rise to. This interpretation is supported by a linear discriminant analysis that is provided with the first 10 principal components of the semantic space. The reduction from 300 to the 10 most important dimensions comes with a 63% drop in accuracy. The semantic structure that underlies the productivity of particle verbs appears to be spread out thinly across all dimensions of the semantic space. This raises the question of what kind of semantics the linear discriminant analysis is detecting.

As a tentative answer to this question, we note that many particle verbs express metaphorical concepts in space and time. For example, the verb *treiben* ‘to drive’ refers to verbs of action, whereas in combination with the particle *unt-* it often takes on the sense of “reducing significance in one’s utterance or report”. For verbs such as *untentreiben*, the particle contributes the metaphorical meaning, but the metaphorical meaning can also be contributed by the base verb. According to the study of verb-particle constructions with *auf-* (Lechler and Roßdeutscher, 2009), even though opaque when considered in isolation, the gist of particle verbs can be remarkably well understood in context. Some particle verbs are entangled with structural metaphors (Lakoff and Johnson, 1980). For instance, *aufblicken*, ‘to look up to somebody’ (from *blicken* ‘to look, glance’) reflects the general structural metaphor that UP IS MORE in the social domain: one looks ‘up’ to someone with ‘higher’ social status. The abovementioned verb *untertreiben*, ‘to understate’, participates in the same structural metaphor, but at the opposite end of the scale. It is also worth noting that particles can realize aspectual meanings. For instance, *aufstehen* (‘stand up’) is telic, whereas its base *stehen* is atelic. Likewise, *aufblasen* ‘to blow up, inflate’ is telic, but its base *blasen* is atelic. A particle verb with *auf* can be quite opaque, and yet realize telic aspect, as in *aufhören*, ‘to stop’, from *hören*, ‘to hear’. Our conjecture is that particles contribute in systematic ways to structural metaphors and to aspect, and that it is these subtle semantic systematicities that the linear discriminant analysis is exploiting. As verbs prefixed with *be-*, *ver-* and *ent-* realize semantics that in many ways resemble those of particle verbs (see Lieber and Baayen, 1993, for an analyses of the corresponding prefixes in Dutch), we expect that they are invisible to t-SNE for the same reasons as the particle verbs.

One quantitative property of particle verbs that is relevant to the present discussion is the negative correlation between $\mathcal{P}(N)$ and $V(N)$. This negative correlation is expected, given the probability theory underlying the $\mathcal{P}(N)$ measure, to arise when each particle corresponds to a set of verbs (of a given size N) sampled from the same underlying population. As the negative correlation between $\mathcal{P}(N)$ and $V(N)$ is modest ($\rho = -0.5$, see Figure 1), words with individual particles are not just random samples from exactly the same population: each particle has its own productivity signature, in line with the results of the linear discriminant analysis given the task to predict particles from the semantic vectors of the particle verbs. Nevertheless, the populations of concepts sampled by individual particles are sufficiently similar for a negative correlation of vocabulary size and vocabulary growth rate to emerge. Since many particles are available for realizing metaphors and aspect, the existence of a negative correlation is unsurprising. For instance, telic aspect can be expressed not only with *auf*, but also with *an* as in *ankommen* (‘arrive’, from *kommen*, ‘come’).

Particles contrast with affixes (except for *be-*, *ver-* and *ent-*) in several remarkable ways. Derived words with suffixes, or with the prefix *un-*, show clear clustering in t-SNE plots. Furthermore, the shift vectors for these affixes also show good clustering. When only the first ten principal components are made available to a linear discriminant analysis, accuracy suffers, not by 63% but only by 23%. Both FRACSS++ and CosClassAvg are better able to predict the semantic vector of the complex word from the semantic vector of its base for affixed words than for particle verbs. The distinct semantic contribution that an affix makes to the meaning of its derived word brings these words close together in semantic space, enabling the t-SNE analysis to detect by-affix clusters in both word space and shift space.

Before reflecting further on the different properties of particle verbs and affixed words that this study documents, we take a step back and consider the reliability of the measurement instrument with which we gauge semantic similarity: word embeddings. For particle verbs, the word embeddings that we used are far from perfect. Particles often occur separated from their base word, and hence are conflated with prepositions by the *fasttext* algorithm. Furthermore, base words that occur separated from their particle may, if the particle is outside the contextual window used by *fasttext*, be conflated with the base word. As a consequence, the amount of imprecision in the word embeddings is necessarily greater for the particle verbs than for the affixed words. However, complex words with the prefixes *be-*, *ver-* and *ent-* also do not cluster in t-SNE plots. Therefore, separability as such is unlikely to be the main cause underlying absence of clustering. In fact, a few particle verbs actually do show clustering in t-SNE plots, even though they are separable. We therefore have good reasons to believe that a major confound with a lack of precision of semantic vectors for particle verbs and their constituents is unlikely to completely drive the present results.

More in general, it should be kept in mind that all embeddings are rife with noise. For instance, English *bank* has two dominant readings, ‘river bank’ and ‘financial bank’. The embedding provided by *fasttext* for *bank* is a frequency-weighted average of the ‘true’ embeddings of river bank and financial bank. (We put true in scare quotes because many words have so many different senses that representing a word with one vector is also widely off even for *bank* in its financial interpretation.) Studies of compounding using distributional semantics wrestle with the same problem. Günther and Marelli (2019); Guenther and Marelli (2022) work with English stand-alone embeddings, just like we do. This, unfortunately, restricts their analyses to compounds that are not written with a space. As a consequence, when their model predicts the embedding of *houseboat* from the embeddings of *house* and *boat*, then the embedding for *house* is confounded with spaced compounds such as *house sparrow*. As spaced compounds are highly productive in English, current work on compound semantics is likely to be biased towards particular semantic relations. In spite of its potential biases, the abovementioned studies on compounding open up new avenues for the study of compound semantics and compound processing. In the same vein, even though the embeddings of particle verbs are noisy, they are not devoid of valuable information. If that were the case, our analyses using LDA as a classifier should have failed to predict particles from the embeddings of the particle verbs, contrary to fact.

To conclude this study, we present some reflections on how the different behavior of particle verbs and affixed words might be understood. As point of departure, we build on a study on productivity (Kastovsky, 1986) which called attention to two distinct functions of morphology, which Kastovsky referred to as ‘labeling’

and ‘syntactic recategorization’. The onomasiological function of word formation is well known: derived words and compounds typically provide labels for things and events that we perceive in the world. But complex words can also be used to refer back to earlier parts of the discourse. For instance, having described a researcher as ‘very thorough’, later on in the text the word ‘thoroughness’ can be used to refer back to this state of affairs (see also [Baayen and Neijt, 1997](#)). For this ‘anaphorical’ use of derivation, it is essential that complex words are transparent and that the affix makes a clear contribution similar to the contribution to meaning made by syntactic modification. For instance, German *aufhören* (on-hear, ‘to stop’) is too unrelated to the meaning of *hören* (‘to hear’) to allow for a construction in which *aufhören* is sensibly used to refer back to an earlier event involving hearing.

German particle verbs seem to us to primarily subserve the labeling function. There are exceptions, for instance, the particles *zurück*, *zusammen* and *durch* show some clustering. But many particles are characterized by large numbers of different senses (e.g., 18 senses for *ab*, and 40 for *an* ([Kempcke, 1965](#); [Kliche, 2009](#)), which can be strikingly different from the meanings of the corresponding prepositions. For instance, *auffallen* ‘to be striking, salient’ and *ausfallen* ‘to fail’ have very different meanings, which in turn are different from the meanings of *fallen* ‘to fall’, *auf* ‘on, up’, and *aus* ‘from’.

Conversely, we have seen that affixed words have meanings that are dominated by the affix: *Absolutheit* (‘absoluteness’) and *Alleinheit* (‘loneliness’) are straightforward nominalizations of the adjectives ‘absolute’ and ‘alone’. Hence affixed words can subserve anaphorical use (Kastovsky’s ‘syntactic recategorization’). Not all affixes have this property, the prefixes *be-*, *ent-*, and *ver-* (but not *un-*) behave more like particle verbs. Of course, affixes often are present in words that serve as labels. Interestingly, ([Baayen and Neijt, 1997](#)) showed, on the basis of a corpus study of Dutch, that the labeling function is typical for the higher-frequency words, and that the anaphoric function is prevalent among the lower-frequency words. Interestingly, the t-SNE maps for affixed words suggest that even the higher-frequency words with a given affix are characterized by senses that are sufficiently related to allow the t-SNE to cluster them reasonably well together.

The particle verbs emerged with quantitative properties that are remarkably similar to those of compound pivots studied by [Shen and Baayen \(2021\)](#) and Shen & Baayen (this volume). Both pivots and particles show a negative correlation for $V(N)$ and $\mathcal{P}(N)$, both show a positive correlation between pivot/particle transparency and $\mathcal{P}(N)$, and neither pivots nor particles show clustering in t-SNE maps. This suggests that particles are closer to compounding than to derivation. In the light of the independent use that particles enjoy, this conclusion is perhaps unsurprising.

As pointed out by [Shen and Baayen \(2021\)](#) for Mandarin adjective-noun compounding, the creation of names for things and events in the world (as perceived by us) can be productive. However, since things and events in the world are “sui generis”, and not necessarily in sync with the categories that evolve in language communities to help make the human-perceived world predictable, a word formation process that creates names runs the risk of not having its own clear semantics. We are thus faced with a transparency paradox: a word formation process that primarily serves the function of providing names for new ideas and concepts that are themselves not compositional, can only be productive if it is semantically relatively unconstrained (see also [Shen and Baayen, 2021](#), and Shen & Baayen, this volume).

This research was made possible in part by funding from stipends from the DAAD and the University of Tübingen (Research & Tübingen) awarded to the first author, as well by funding from the ERC (project WIDE-742545) awarded to the second author.

References

- Aronoff, M. (1976). *Word Formation in Generative Grammar*. MIT Press, Cambridge, Mass.
- Baayen, R. H. (1993). On frequency, transparency, and productivity. In Booij, G. E. and van Marle, J., editors, *Yearbook of Morphology 1992*, pages 181–208. Kluwer Academic Publishers, Dordrecht.
- Baayen, R. H. (2001). *Word Frequency Distributions*. Kluwer Academic Publishers, Dordrecht.
- Baayen, R. H. (2005). Data mining at the intersection of psychology and linguistics. In Cutler, A., editor, *Twenty-first century psycholinguistics: Four cornerstones*, pages 69–83. Erlbaum, Hillsdale, New Jersey.
- Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*.
- Baayen, R. H. and Lieber, R. (1991). Productivity and English derivation: a corpus-based study. *Linguistics*, 29:801–843.
- Baayen, R. H. and Neijt, A. (1997). Productivity in context: a case study of a Dutch suffix. *Linguistics*, 35:565–587.
- Baayen, R. H. and Renouf, A. (1996). Chronicling The Times: Productive Lexical Innovations in an English Newspaper. *Language*, 72:69–96.
- Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.
- Bonami, O. and Paperno, D. (2018). Inflection vs. derivation in a distributional vector space. *Lingue e Linguaggio*, 17(2):173–195.
- Corbin, D. (1987). *Morphologie derivationelle et structuration du lexique*. Niemeyer, Tübingen.
- Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, 40:237–264.
- Guenther, F. and Marelli, M. (2022). Patterns in caoss: Distributed representations predict variation in relational interpretations for familiar and novel compound words. *Cognitive Psychology*, 134:101471.
- Günther, F. and Marelli, M. (2019). Enter sandman: Compound processing and semantic transparency in a compositional perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(10):1872.
- Kastovsky, D. (1986). Productivity in word formation. *Linguistics*, 24:585–600.
- Kempcke, G. (1965). *Die Bedeutungsgruppen der verbalen Kompositionspartikeln an-und auf-in synchronischer und diachronischer Sicht. Beiträge zur Geschichte der deutschen Sprache und Literatur, volume 87*.
- Kliche, F. (2009). *Zur Semantik der Partikelverben auf ab. Eine Studie im Rahmen der Diskursepräsentationstheorie*. PhD thesis, Master’s thesis, Universität Tübingen.
- Krijthe, J. H. (2015). *Rtsne: T-Distributed Stochastic Neighbor Embedding using Barnes-Hut Implementation*. R package version 0.15.
- Lakoff, G. and Johnson, M. (1980). *Metaphors we live by*, volume 111. Chicago London.
- Landauer, T. and Dumais, S. (1997). A solution to Plato’s problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104(2):211–240.

- Lechler, A. and Roßdeutscher, A. (2009). Analysing german verb-particle constructions with 'auf' within a drt based framework.
- Lieber, R. (2010). *Introducing Morphology*. Cambridge University Press, Cambridge, UK.
- Lieber, R. and Baayen, R. H. (1993). Verbal prefixes in Dutch: a study in lexical conceptual structure. In Booij, G. E. and Marle, J. v., editors, *Yearbook of Morphology 1993*, pages 51–78. Kluwer Academic Publishers, Dordrecht.
- Maaten, L. v. d. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605.
- Marelli, M. and Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review*, 122(3):485.
- Marelli, M., Gagné, C. L., and Spalding, T. L. (2017). Compounding as abstract operation in semantic space: Investigating relational effects through a large-scale, data-driven computational model. *Cognition*, 166:207–224.
- Riddle, E. (1985). A historical perspective on the productivity of the suffixes *-ness* and *-ity*. In Fisiak, J., editor, *Historical Semantics, Historical Word-Formation*, pages 435–461. Mouton, New York.
- Schreuder, R. and Baayen, R. H. (1994). Prefix-stripping re-revisited. *Journal of Memory and Language*, 33:357–375.
- Shafaei-Bajestan, E., Moradipour-Tari, M., Uhrig, P., and Baayen, R. H. (2022). Semantic properties of English nominal pluralization: Insights from word embeddings. *arXiv arxiv.org/abs/2203.15424v1*.
- Shahmohammadi, H., Lensch, H., and Baayen, R. H. (2021). Learning zero-shot multifaceted visually grounded word embeddings via multi-task training. *CoNLL 2021*. arXiv preprint arXiv:2104.07500.
- Shen, T. and Baayen, R. H. (2021). Adjective–noun compounds in mandarin: a study on productivity. *Corpus Linguistics and Linguistic Theory*.