The role of meaning in the rivalry of *-ity* and *-ness*: evidence from distributional semantics

Abstract

Both *-ity* and *-ness* are frequent and productive suffixes in English that fulfill the same core function: turning adjectives into nouns that denote the state or quality of whatever the adjective denotes. This well-known affix rivalry raises two core questions: (1) What determines the choice between *-ity* and *-ness* for a given base word? (2) Are the two affixes synonymous? For the first question, previous work has focused on morphological and phonological properties of the bases, but not their semantics. For question 2, the literature fails to give a convincing answer, with some studies, faced with doublets like *ethnicity/ethnicness*, arguing for a semantic difference, but most assuming synonymy. Using pretrained distributional vectors, I show empirically that (1) the semantics of the bases plays a major role in affix selection and (2) the two affixes induce similar meaning shifts.

1 Introduction

The two suffixes *-ity* and *-ness* are both very frequent and, on the face of it, seem to fulfill exactly the same core function: deriving a noun describing a quality or state from an adjective. The noun *redness* describes the quality or state of being red, and the noun *insularity* the quality or state of being insular. But although in these two cases the addition of the suffix seems to have the same effect, they are also a good example for the different behavior of the two suffixes: when searching in a huge up-to-date corpus like iWeb (Davies, 2018), one does neither find *redity* nor *insularness*, illustrating that the two suffixes do not seem to be attracted to the same adjectival bases.

At the same time, it is easy to find bases that occur with either affix, sometimes with a clear meaning difference between the two resulting derivations on specific usages, cf. *activity* vs. *activeness* in (1) and (2).

(1) Can I still get together with my friends and have coffee, play golf, go bowling or any **activity** that I still enjoy doing? [iWeb]

(2) Such different charecteristics [sic] as calmness and **activeness** are harmoniously combined in me. [iWeb]

But it is similarly easy to find instances with no such obvious discernable meaning difference, cf. *inclusivity* in (3) vs. *inclusiveness* in (4).

(3) And part of me believes that this **inclusivity** of calling us the LGBTQQTY-whatever-LMNOP tends to stress our differences. [iWeb]

(4) The University will create an environment that promotes diversity through a culture of civility and **inclusiveness**. [iWeb]

These observations drive the two questions addressed in the paper: what exactly determines the distribution of affixes with respect to bases, and what, if any, are the meaning differences between the two suffixes? While the first question has been much discussed and many pertinent factors have been identified, the role of the semantics of the bases in this has not been considered on a quantitative empirical basis. For the second question, there have been several attempts at identifying a semantic difference, but no large scale empirical evidence has been put forward.

I use distributional semantics to address both questions, hypothesizing that: (1a) If the semantics of the bases drives affix selection, a clear semantic difference between bases taking *-ity* and bases taking *-ness* is expected. (1b) This difference should even hold for complex bases that end in the same suffix (e.g. *-ive*). These cases are of particular interest, because selection of either *-ity* or *-ness* cannot be based on the form of the ending or the final morpheme itself, as it stays the same. (2a) If *-ity* and *-ness* are synonymous, the change in meaning they bring about should be the same; in terms of distributional semantics, the shift in semantic space induced by the two suffixes should be the same for both *-ity* and *-ness* derivatives. (2b) Doublets (such as *aggressivity–aggressiveness*) should show no systematic difference in the semantic vectorspace.

2 Background and objectives

While both suffixes occur also with other bases, especially nouns (see Bauer, Lieber, and Plag 2013, chapter 13 for an overview), I will restrict my analysis to *-ity* and *-ness* derivatives with adjectival bases. This approach abstracts away from the complications arising with prepositions, nouns or phrases as bases (see Bauer, Lieber, and Plag 2013 for discussion). I will leave it to future research to explore the semantic space of these derivatives.

2.1 The role of the base in affix selection

Historically, *-ness* is the older of the two suffixes. It has been used since Old English to turn adjectives into nouns. Perhaps as a reflection of this, the majority of its bases are native although the suffix "may be tacked on to any adjective" (Marchand 1969, p. 335). In contrast, *-ity* is newer, with *-ity* words, being whole word borrowings from French, starting in the 14th and 15th century (cf. Riddle 1985 and Lindsay 2012 for the complex borrowing history). This history is in line with the often observed aversion of *-ity* to native bases (Marchand 1969, p. 314) or even the claim that it is restricted to just latinate bases (Aronoff 1976, p. 51). Leaving history aside, there are intriguing differences in the distribution of the two affixes even in present day English. On hapax-conditioned productivity measures, *-ness* is more productive than *-ity* (cf. Baayen and Renouf 1996, Plag 2006). Lindsay (2012), based on google hits for 3256 potential rival pairs, arrives at a more differentiated picture when looking more closely at potential bases in terms of their endings: while *-ness* is overall more productive (in terms of its distribution across bases), *-ity* dominates in some subdomains. For example, bases ending in *-ing*, *-ish* and *-ful* occur only with *-ness* in Lindsay's data, but both suffixes occur to a considerable extent with bases ending in *-ous/-os* and *-ive*, and *-ity* is dominant for bases ending in *-able, -al, -ic,* and *-ar* (see also Anshen and Aronoff 1981 for experimental support of an *-ity* preference for *-able/-ible* bases).

Lindsay (2012) discusses this distributional difference in terms of a morphological constraint on *-ity*, but Arndt-Lappe (2014) points out that previous studies do not allow for a distinction between a preference for either suffix based on the morphological makeup of the base, or just the form of the base. To take just one example, is the decisive feature for the status of the adjective *affective* as a potential base for *-ity* or *-ness* just its form, that is, its ending on the string *-ive*, or that it consists of a base and the morpheme *-ive*? In her own analogical modeling of whether a base takes *-ity* or *-ness*, Arndt-Lappe (2014) used a mostly form-based coding, with word-status (word vs. non-word, e.g. phrase/bound form) the only non-form-based information available to the model. The form-based coding was constituted by six phonetic features describing the two final syllables of the base. Using the set of the 564 twentieth century neologisms in the OED as both training set and test set with a leave-one-out approach, her simulation yields a macro-averaged F-score of 0.88 (see section 3.3 for more detail on the F1 or F score and how it is calculated), indicating that phonetic makeup of the last two syllables and the word-status of the base are good predictors of whether a neologism will end in *-ity* or *-ness*. This result suggests that the morphological status of the endings themselves is irrelevant.

While the morphological make-up of the base and its form features are central to the discussion in the literature, Riddle (1985) points out that there are some semantic groups of bases that show very clear preferences, regardless of form. For example, color words only go with *-ness*. In addition, Riddle (1985) argues that the suffixes within the bases might actually influence the choice between *-ity* and *-ness* because the suffixes still have semantic significance. In other words, that bases ending on *-ful* prefer *-ness* might equally well be based on a preference of *-ness* for a form feature, or for the specific morpheme *-ful*, or for adjectives whose meanings are similar because they all are formed with the help of this morpheme.

As we will shortly see, distributional semantics allows us to explore the role of semantics in affix selection on a large empirical scale.

2.2 Meaning differences between *-ity* and *-ness*

Exclusively looking at features of the base as driving force in the selection of either *-ity* or *-ness* only makes sense under the assumption that *-ity* and *-ness* are synonyms, an assumption that is often left implicit and arises only from meaning descriptions. For example, Marchand (1969, p. 312) characterizes *-ity* as “form[ing] abstract substantives from adjectives with the meaning ‘state, quality, condition of –’”, and *-ness* as “form[ing] abstract substantives with the meaning ‘state, quality, condition of –’” Marchand (1969, p. 334). For him, the only difference between the two suffixes lies in the restriction to adjectival bases for *-ity*, otherwise they are exactly synonymous. Riddle (1985, p. 437) markedly breaks with this assumption of synonymy, claiming that “the suffixes themselves have different meanings when occurring on many bases, but the distinction is not realized on all bases”. To support her view, she discusses a number of doublets, that is, pairs of *-ity* and *-ness* derivatives from the same base like *hyperactivity* and *hyperactiveness*, and argues that

> “*-ness* tends to denote an embodied attribute or trait, while *-ity* tends to denote an abstract or concrete entity. Examples of what I consider to be abstract entities are the names of concepts and situations and of characteristics in the generic sense. For example, an *-ity* word may refer to a characteristic, in the generic sense, while there is a tendency for the corresponding *-ness* word formed on the same base to describe an embodied attribute.” Riddle

(1985, p. 437).

Supporting evidence for her are the usages of *-ity* and *-ness* in the two passages drawn from the same 1982 newspaper article, her (1):

(5) a. "However, don't call this third-grader a picky eater. She's a selective one, a Feingold diet subscriber, whose *hyperactiveness* has decreased, her mother says, since she began the program four years ago."

b. "But to date there is no evidence that this type of dietary regime will have any effect on *hyperactivity* in children."

In (5-a), *hyperactiveness*, so Riddle, "denotes an embodied attribute of a particular child" (438), while *hyperactivity* in (5-b) denotes the condition. Riddle's distinction has received much comment. Cowie (1999, p. 263) convincingly argues that it is simply not possible to consistently distinguish between Riddle's attribute and abstract entity senses. Bauer, Lieber, and Plag (2013, p. 257) understand this difference in terms of the reification of the quality denoted by the adjective, that is, here, *hyperactivity* is reified because it is the name of a diagnosable condition. However, they note that this kind of difference does not obtain across the board for all doublets, and hypothesize that the observed difference might be linked to

> "the greater propensity of forms in *-ity* to be high frequency established forms and to have lexicalized meanings. Lexicalized forms can denote reified concepts or concrete objects. And indeed many *-ity* forms have such reified denotations. Indicative of this reification is the fact that some *-ity* nouns have become count nouns on specific readings, [...] " Bauer, Lieber, and Plag (2013, p. 257).

Baeskow (2012) is likewise dissatisfied with Riddle's description of the nature of the observed difference and instead argues that *-ness* is sensitive to the scalar structure of its base, tending to select a 'large degree of ADJ' reading. One effect of this reading is the availability of degree interpretations even if the base adjective is non-gradable, e.g. in *intense aliveness*. In contrast, *-ity* selects the bare property, which is, according to Baeskow, in line with its preferred usage with latinate scientific terms, for example the *-ic* and *-il* bases. Perhaps most interestingly, Baeskow (2012, pp. 26–30), follows up on a footnote in Riddle (1985) suggesting a link between between *-ness* and *-ity* usages and specific and generic readings. Baeskow argues that there is indeed such a connection between the two affixes and the possible readings. This preference for different readings is in turn linked to different syntactic environments: specific usages often occur together with an explicit realization of the external argument of the base adjective. E.g., *the pureness of the water* receives a specific reading already because of the *of*-PP: we are talking about the specific pureness of a specific water. Even so, Baeskow (2012) is careful to note that both affixes may be used specifically as well as generically, and the exact quantitative basis of her observations does not become clear. The counterexamples that Bauer, Lieber, and Plag (2013) adduce against Riddle's analysis are similarly unexpected on Baeskow's account. Riddle's and Baeskow's accounts suffer from the fact that even for the doublets neither explanation works satisfactorily, and that for the majority of derivatives, which do not have a corresponding form with the other suffix (from now on called 'non-doublets' in this paper), the proposed semantic difference of the two affixes seems to play a role only for a very limited subset of bases. Again, distributional semantics allows us to investigate this on a large empirical scale.

2.3 Other aspects in the distribution of *-ity* and *-ness*

Two other aspects in the distribution of *-ity* and *-ness* derivatives that have received considerable attention are register and gender.

Plag, Dalton-Puffer, and Baayen (1999) lament that "very little attention has been devoted to the role derivational morphology may play in register variation", pointing out that nominalizations in *-tion, -ment, -ness* or *-ity* are the only clearly word-formation related feature of the overall 67 linguistic features used in Biber (1988) for the analysis of register in English. Since *-ity* and *-ness* are both part of the same feature set, they cannot themselves distinguish between registers there. They are part of one feature set, because they both serve the function of compressing information into more compact form and “promoting” more abstract concepts Biber (1986, p. 395). Plag, Dalton-Puffer, and Baayen (1999) further observe that they play only a limited role in distinguishing between the six basic factors of variation in English as identified in Biber (1988) and further analyzed in Biber (1995): nominalizations are associated with elaborated reference and emerge as one of the significant factors only for Biber’s dimension 3, situation-dependent versus elaborated reference. Plag, Dalton-Puffer, and Baayen (1999) themselves analyze both *-ity* and *-ness* individually, along with a third suffix producing abstract nominals, *-ion*, and 12 other derivational affixes. They investigate differences in the productivity of the individual suffixes across speech and writing by using the three domains written, spoken context-governed, and spoken demographic language of the British National Corpus (BNC, 2007). While they find that suffixes usually differ in their productivity across registers, they also find that these patterns are not uniform. In terms of differences between *-ity* and *-ness* in particular, they report that *-ness* is more likely to be used to coin new words, but, importantly, this difference obtains across all three registers. In terms of the average number of types, *-ity* and *-ness* appear to behave relatively similar.

Clearer differences in *-ity* and *-ness* use are, in contrast, reported in works using historical data. Rodríguez-Puente, Säily, and Suomela (2022) show that *-ity* became more frequently used relative to *-ness* during the Early Modern English Period, and that this development was linked to register in that it started in written registers and only later spread to speech-related registers (e.g., ranging from diaries and letters to drama and sermons). Cowie (1999) investigates both affixes in the ARCHER corpus ranging from 1650 to 1990 and finds different register preferences, for *-ity* in scientific and medical writing, for *-ness* in sermons and fiction, but constating that “all registers have derivations in both suffixes, and there is no clear and unambiguous preference” (Cowie 1999, p. 248).

For gender, Säily and Suomela, 2009 found no difference for *-ness*, but a clear difference for *-ity* in the 17th century part of the Corpus of Early English Correspondence: it is used less productively by women. This difference, however, disappears in the 18th century part of the same corpus Säily (2018). For Present Day English, Säily (2011), using the BNC, found again clear gender differences. For *-ity* in written English, the results are the same as for the 17th century: women use *-ity* less productively. In contrast, the same diverse usage is found for *-ness* across both genders. In the spoken-subcorpus, both suffixes are used less productively by women, with the difference tied to lower-class women in the case of *-ness*.

2.4 Distributional semantics in morphology

Over the last decade, distributional semantics has become a mainstream method in linguistics (Boleda, 2020), including word-formation semantics (cf. the overview and the three papers by Kotowski and Schäfer, Bonami and Naranjo, and Schäfer in Kotowski and Plag 2023). The fundamental idea behind distributional semantics is that the meaning of a word is reflected in its distribution, with the distributional hypothesis that "[w]ords with similar distributional properties have similar meanings" Sahlgren (2006, p. 21). Key to the computational implementation of this idea is the step of encoding the distribution of a word by means of a vector in geometrical space.

To see how distributional semantics works and how it can fruitfully be used for morphological questions, let's say we are interested in whether the two derivatives *inclusivity* and *inclusiveness* are semantically more or less similar to each other than to their base, *inclusive*. For our toy example of a distributional semantics analysis, we collect the cooccurrences of *inclusivity*, *inclusiveness*, and *inclusive* with different contexts words, for example the nouns *environment*, *part*, or *price*. Tabulating the cooccurrences gives us table 1, where the first cell tells us that *inclusive* cooccurred 2 times with *environment*.

We now take the cooccurrence counts to represent vectors, so that the vector (2,1,4) represents *inclusive*, and the vector (4,3,1) represents the word *inclusiveness*, that is, the vectors correspond one-to-one to the rows in the tabulated cooccurrences. Figure 1 illus-

Table 1: Toy example illustrating the first step in creating a simple distributional model: collecting cooccurrence counts for the target words, here the cooccurrences of *inclusive*, *inclusiveness*, and *inclusivity* with the three nouns *environment*, *part*, and *price*.

	environment	part	price	column total
inclusive	2	1	4	7
inclusiveness	4	3	1	8
inclusivity	3	4	1	8
row total	9	8	6	23

trates how these vectors can be mapped into geometrical space: the context words, here *environment*, *part*, and *price*, represent the dimensions of the vectorspace, and the cooccurrence counts with the target words determine the length and directionality of the three vectors, each of which encodes the distribution of the corresponding lexeme.

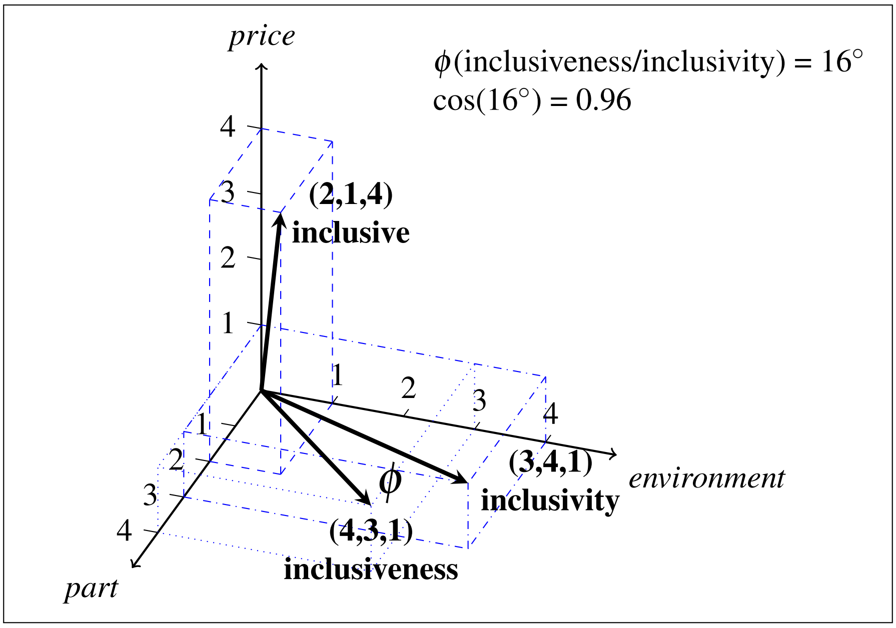

Figure 1: A three-dimensional space showing the distributional vectors of *inclusiveness*, *inclusivity* and *inclusive* based on the toy data in table 1. The three dimensions stand for the cooccurrences with the nouns verbs *environment*, *part*, and *price*.

This visualization by itself already shows that the three vectors are not equally similar to each other. Among the three, the vector representing *inclusive* is clearly the outlier, and we already can gauge the answer to our initial question from the visualization: apparently, the two derivatives are more similar to each other than to their base. A very common measure to quantify similarity between vectors is the cosine similarity (Boleda, 2020).

The cosine similarity is simply the cosine of the angle that holds between any two vectors. The angle ϕ between the vector for *inclusiveness* and the vector for *inclusivity* in figure 1 is 16°, taking the cosine yields the cosine similarity of 0.96. In contrast, the vectors of the two other pairings, *inclusiveness* and *inclusive* as well ass *inclusivity* and *inclusive*, are far less similar on this measure. For the first pair, the 50° angle results in a cosine similarity of 0.64, for the second pair, the angle of 53° results in a cosine similarity of 0.60.

The cosine similarity has been established as a good stand-in for semantic similarity between the words represented by the vectors, with correlations between human judgments of semantic similarity between word pairs and the corresponding cosine similarity commonly used as benchmarks for the quality of distributional models (Baroni, Dinu, and Kruszewski, 2014). The highest possible cosine similarity value is 1, the angle between the two vectors is 0 degree and the distribution as captured by the vectors is extremely similar (and in fact identical if all vectors have the same length). In contrast, a cosine similarity value close to zero indicates that the vectors are separated by a 90 degree angle and are basically unrelated. In our toy example, the cosine similarities suggests that *inclusiveness* and *inclusivity* are extremely semantically similar, while cosine similarities between the vector for *inclusive* and the two derivatives suggest that it is less semantically similar to both.

The vectors in our toy example can be visualized easily because we just need three dimensions. To visually explore the high-dimensional vectors that are standardly used in distributional semantics, dimension reduction-techniques can be applied, usually mapping the high-dimensional vectors into two-dimensional space. This paper uses the t-SNE dimension-reduction-technique and and subsequent visualization, which is explained in more detail in section 3.3.

Finally, classification techniques can be applied to check for patterns in the vectors. This paper use Linear Discriminant Analysis (LDA) for this purpose. LDA is a supervised classification method that predicts the class of an item, in this case a word, out of a set of given classes with the help of numerical predictors, in this case, the corresponding word vectors. For our toy example, we could check whether LDA can correctly classify the three words into the two classes BASE and DERIVATIVE based on their vectors. Since LDA lets us quantitatively express the success of its application, it is the ideal complement to the combination of t-SNE and subsequent visualization. Again, this is explained in more detail in section 3.3.

While the first distributional semantic encodings were based on cooccurrence counts and additional normalization steps, since the publication of the word2vec algorithm (Mikolov et al., 2013), word vectors are most often created via machine learning, and many thus trained vector sets are freely available. In a comparative study by Baroni, Dinu, and Kruszewski (2014), these vectors, often called embeddings, on average outperform the count models. The set of vectors, fastText (Mikolov et al., 2017), used in this paper is such a pre-trained vector set, in which the word vectors have 300 dimensions (see section 3.3 for further details).

Before closing this section on distributional semantics and morphology, one word on the status of distributional vectors with regard to the possible difference between *-ity* and *-ness* derivatives due to their usage in different registers or by different speakers. Both types of differences, if they are indeed reflected in different distributions, can and ideally will contribute to differences between the vectors of the respective words. I will come back to the consequences of this for the interpretation of the results of this investigation in the respective discussion sections.

2.5 Expectations and hypotheses

I will address the following research questions: (1) What is the role of the base adjectives' semantics in determining the choice between *-ity* and *-ness* and (2) Are the two suffixes synonymous? Concerning the role of base semantics in affix selection, I hypothesize that:

(1a) If the semantics of the bases drives affix selection, a clear semantic difference between bases taking *-ity* and bases taking *-ness* is expected.

(1b) This difference should even hold for complex bases that end in the same suffix (e.g. *-ive*), where neither form-based features nor the final morpheme itself can be used to determine affix choice.

To address the issue of synonymy of *-ity* and *-ness*, I hypothesize that:

(2a) If *-ity* and *-ness* are synonymous, the change in meaning induced by the two suffixes should be the same for both *-ity* derivatives and *-ness* derivatives.

(2b) Doublets (such as *aggressivity/aggressiveness*) should show no systematic meaning difference.

These questions are addressed via two separate studies. The first study investigates non-doublets with a dataset of *-ity* derivatives and their bases and *-ness* derivatives and their bases. The second study investigates the semantics of doublets. In both studies, I also consider whether the absolute frequency of the derivatives plays a role, hypothesizing that higher frequencies, taken as indicators of lexicalization, might also be associated with a higher proportion of idiosyncratic meaning shifts that might obscure an understanding of the typical behavior of the derivatives relative to their bases. This is particularly relevant for the second study, since lexicalization has been considered as a factor for non-synonymy of *-ity*/*-ness* doublets.

For both study 1 and study 2, all preparatory steps and all further analysis are reproducible via the scripts in `https://figshare.com/s/c1e2d11fa1b6825bf777` [THIS IS A PRIVATE LINK. DO NOT SHARE PRIOR TO PUBLICATION].

3 Study 1: semantic clustering of bases and derivatives

The first study focuses on hypotheses (1a), (1b), and (2a). To address whether the semantics of the bases is linked to a preference for *-ity* or *-ness*, I test whether non-doublet bases, that is, bases that occur only with either one of the two affixes in the dataset, are semantically different, using distributional vectors as representations of word meanings.

To investigate the semantics of the *-ity* and *-ness* affixes for the non-doublets, we compare the clustering of base vectors with regard to their selected affix to the clustering of the derivative vectors themselves. To address (2a) in more detail, I also investigate the cosine similarities between bases and derivatives, and the possible influence of the lemma frequencies of both bases and derivatives on these.

3.1 Materials and techniques

I first sampled all pairs of adjectival bases and *-ity/-ness* derivatives in the ukWaC corpus (Baroni et al., 2009) with pretrained fastText vectors (Mikolov et al., 2017). These vectors are analyzed with t-SNE, a dimensionality reduction technique, and Latent Discriminant Analysis, a classification technique. For comparisons of the cosine similarities, standard statistical comparisons and regression analyses are used.

3.2 Dataset

The ukWaC corpus is a web-derived 2 billion word corpus of English (see Baroni et al. 2009). Its size is big enough to contain enough lower frequency *-ity* and *-ness* words, while at the same time not too big to be used outside of high-performance computing contexts. It is fully part-of-speech-tagged and lemmatized with TreeTagger (`https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/`), and conveniently provides frequency lists. The ukWaC corpus does not come with pretrained vectors, nor are pretrained vectors based on it available elsewhere. In principle, it is possible to derive vectors from the ukWaC, but this would require computing power and subsequent evaluation of the quality of the vectorspace that far exceeds the bounds of what is possible and reasonable to do in this paper. In addition, given that it is well-known that vectors created for low-frequency words tend to be unreliable, with 50 occurrences often used as a threshold for the minimal occurrence of one lemma, using the ukWaC itself as the base for the vectors would also have meant the exclusion of almost half of the pairs that were selected in the ukWaC for the investigation (see below for the exact selection procedure).

Instead, I looked for a pretrained vectorset that has been trained on language data that is by and large comparable to the basic of the ukWaC, and that, in addition, contains vectors for a high number of base-derivative pairs in the ukWaC. The vectorset that best fulfilled these conditions were the pretrained fastText vectors (Mikolov et al., 2017).

The fasttest vectors I used, `wiki-news-300d-1M.zip` available at `https://fasttext.cc/docs/en/english-vectors.html` [last access 2024-03-18], are trained on 16 billion tokens, using a corpus that itself was concatenated from 5 different web-sources and web-derived corpora, see Mikolov et al. (2017) for the details. I chose the variant of fastText that does not include subword information, as this subword information would automatically help to differentiate between our target derivatives via their different endings, *-ity* and *-ness*. This raw corpus is not available (and can not easily be reconstructed). Importantly, for our purposes it is similar enough to the corpus base of the ukWaC in that exclusively English-language web-based content is used. One additional advantage of using a set of independently trained vectors is the very fact that it was created independently from the main questions of this paper, that is, the vectorspace has not been created with the aim of maximally distinguishing between *-ness* and *-ity* items. That the corpus base is eight times the size of the ukWaC corpus guarantees that this vectorset will provide a sufficiently large number of vectors of *-ity* and *-ness* derivatives and their bases.

The combination of the ukWaC corpus and the fastText vectors thus guarantees a large number of vector presentations for base-derivative pairs and allows us to meaningfully include part-of-speech tagging and lemma frequencies in our further data preparation and analysis. To establish the pairs of base adjective and derived *-ity* or *-ness* words, the following steps were taken:

A: Identifying all base-derivative pairs in the ukWaC.

1. To obtain the derivatives, I selected all forms ending in either *-ity* or *-ness* from the ukWaC unigram lemma lists. These lists contain 33011 *-ity* and 17796 *-ness* items, with upper and lowercase items treated as distinct.

2. These lists were first cleaned by keeping only items consisting of alphabetic characters only, or of alphabetic characters in combination with a single non-initial hyphen. Further, lower- and uppercase lemmata were merged. Some compounds with high frequent derivatives as final element were also excluded (for example, hyphenated compounds ending in *-capacity* or *-business*). This leaves 20509 *-ity* lemmata and 15464 *-ness* lemmata.

3. Adding possible bases to each item in the list and filtering
   Both *-ity* and *-ness* show variation in the way the adjectival base is combined with the affix. To find an adjectival base for a given derivative, possible base forms were created by exploiting the following patterns, each exemplified by an existing base/derivative pair:

(6) a. Base/Derivative variation patterns for *-ity*:
   (i) `ble/bility`: *possible/possibility*
   (ii) `ous/osity;ocity/ocious|ous` *curiosity/curious*; *atrocity/atrocious*
   (iii) `que/city;cious/city`: *opaque/opacity*; *audacious/audacity*
   (iv) `e/ity`: *active/activity*
   (v) no adaption: *brutal/brutality*

   b. Base/Derivative variation patterns for *-ness*:
   (i) `y/i`: *happy/happiness*
   (ii) no adaption: *effective/effectiveness*

Whenever a possible base was added, the resulting pair was only kept when the possible base also occurred in the ukWaC unigram list. If not, the next orthographic variant was considered, until a base occurring in the unigram list had been found or all possibilities had been exhausted, in which case the derivative was not further considered. This filtering followed the ordering of variation in (6). To illustrate, for *brutality*, the first three patterns, (6-a-i)–(6-a-iii), do not match the ending and were therefore skipped. The next pattern fits, but the resulting possible base, *brutale*, does not occur in the unigram list, therefore the final pattern, just stripping the *-ity*, is used, successfully identifying the pair *brutal*-*brutality*.

This results in 7285 *-ity* base-derivative pairs and 8785 *-ness* base-derivative pairs.

B: Final filtering steps

1. The resulting items were checked against all available fastText-vectors, and only those pairs were kept for which the vectorspace provided both a vector for the derivative as well as for the base. This results in 1572 *-ity* pairs and 1836 *-ness* pairs.

2. Only those pairs were selected for which the base was tagged as an adjective in the ukWaC (including cases where this was not the only POS-tag). Some items with obvious spelling errors or nonstandard spelling variants were also excluded (e.g. *possiblity* etc.). This leaves 1475 *-ity* pairs and 1802 *-ness* pairs.

No attempt was made to identify or exclude items that might be construed via different pathways. For example, *unchastity* is here paired with *unchaste* because it fits the *-e/-ity* pattern and the adjective *unchaste* exists in the ukWaC corpus. Whether a derivation via *un-* prefixation from *chastity* is historically or psychologically more adequate was not explored: neither is there the relevant etymological data for all pairs, nor are there any clear criteria to decide when etymologically correct derivations do or do not correspond to the actual synchronic analyses of speakers.

From this dataset I eliminated the doublets and set them aside for separate analysis (see study 2). This leaves us with 1343 *-ity* and 1671 *-ness* pairs. The derivatives cover a wide frequency range for both *-ity* and *-ness* words. Table 2 shows the distribution of the derivatives across the full frequency range by binning the lemmata via their frequencies. The bins were chosen to give a representative overview across the frequency ranges, keeping an eye on an even distribution of the two affixes in each bin and reasonable bin sizes.

Table 2: Distribution of the *-ity* and *-ness* lemmata in terms of token frequency. Doublets are excluded.

frequency range	freqBand	*-ity* lemmata	*-ness* lemmata
01 – 09	ultraLow	192	237
10 – 49	low	353	690
50 – 149	mid	250	346
150 – 499	midHigh	197	197
500 – 1999	high	139	135
2000 – 9999	superHigh	115	48
10 000 – ...	ultraHigh	97	18

The distribution of *-ity* and *-ness* derivatives in terms of token frequency is in line with the overall greater productivity of *-ness*, reflected in more *-ness* derivatives in the lower frequency bins. It is also in line with Bauer, Lieber, and Plag (2013, p. 257) in that the distribution clearly shows a propensity of *-ity* derivatives in the higher frequency ranges.

Table 3 shows the distribution of the *-ity* and *-ness* lemmata over notable patterns in their end-strings, exemplified with examples from the data. Classification is exclusively based on orthography. Patterns were manually identified by sorting all bases by their reverse endings and looking for repeated characters and morphemes.

Again, this distribution is in line with previous studies. For example, that *-ble, -al, -ic, -ar* favor *-ity* and *-ed, -ing, -less, -ish, -ful, -nt, -ous, -ive* favor *-ness* also emerges in Lindsay (2012). The table also shows that form-based and meaning-based factors in the choice between *-ity* and *-ness* for bases that fall in notable form classes often cannot be teased apart because the distribution within specific forms is too uneven. Only bases ending in the string *-ive* have a sufficiently high number of both *-ity* and *-ness* derivatives (90 *-ity* and 108 *-ness*). This subset, that is, the non-doublets with bases ending on *-ive*, will be analyzed separately because it allows to investigate the influence of meaning independent of form.

Table 3: Distribution of endings in the adjective bases of *-ity*/*-ness* derivatives, with one example for each affix. Doublets are excluded.

base ending	*-ity* lemmata	*-ness* lemmata	example *-ity*	example *-ness*
al	291	3	*sentimentality*	*gradualness*
ar	55	5	*insularity*	*dearness*
ble	547	10	*responsibility*	*nimbleness*
ed	0	173		*learnedness*
ful	0	83		*playfulness*
ic	120	3	*hapticity*	*epicness*
id	31	7	*validity*	*horridness*
ile	29	1	*sterility*	*vileness*
ing	0	20		*amazingness*
ish	0	52		*stylishness*
ive	90	108	*narrativity*	*distinctiveness*
ld	0	7		*coldness*
less	0	97		*cluelessness*
ly	0	82		*deadliness*
nd	3	17	*fecundity*	*blandness*
nt	4	14	*quantity*	*pleasantness*
ous	19	163	*sagacity*	*precariousness*
te	1	46	*unchastity*	*cuteness*
y	0	384	*seediness*	
OTHER	153	396	*complexity*	*slackness*

3.3 Analysis

The main analysis of the data consists of two steps that are explained in more detail below. First, the 300-dimensional word vectors are mapped onto two-dimensional space using t-Distributed Stochastic Neighbor Embedding (t-SNE), a technique designed to reveal patterns in high-dimensional data. These patterns can be identified directly from the visualization. It is an unsupervised, exploratory approach to the data. Second, I use a classification technique, Linear Discriminant Analysis (LDA), to corroborate the results of the visual inspection of the mapping. This, in contrast, is a supervised approach: the technique is used to find the best way to classify the data in the classes identified by me beforehand. With this two-step approach, I follow Shafaei-Bajestan et al. (2022). For the analysis of the cosine similarities, I use standard inferential statistics and regression modeling.

t-Distributed Stochastic Neighbor Embedding (t-SNE) (Maaten and Hinton 2008), is a dimension reduction technique that can map high-dimensional data on just two dimensions. To do so, the high-dimensional data is first converted into a matrix of pairwise similarities, and the aim of the of t-SNE is to efficiently find a low-dimensional equivalent that is as faithful as possible to the pairwise similarities. Searching for a transformation that minimizes the error between input and output similarities, "t-SNE is capable of capturing much of the local structure of the high-dimensional data very well, while also revealing global structure such as the presence of clusters at several scales" Maaten and Hinton 2008, p. 2587. Visualizing the resulting dimensions has proven highly successful in cluster-detection (Arora, Hu, and Kothari 2018; see also Maaten and Hinton 2008 for comparison with other dimensionality reduction techniques on a variety of datasets). My usage of this technique for the analysis of the word vectors is inspired by Shafaei-Bajestan et al. (2022), who used it to analyze the semantic properties of English nominal pluralization. In using t-SNE, a number of parameters can be set that influence the behavior of the algorithm. The settings used in this study are the same as used in Shafaei-Bajestan et al. (2022). This was done because their study was also on English morphology, and, more importantly, the results for my data when using these settings showed clear and interpretable patterns. Note that this does not mean that other settings would not have produced even better interpretable results, the interested reader can experiment with the settings when inspecting the data and scripts that come with this paper where the detailed setting are documented.

Linear Discriminant Analysis (LDA), in contrast, is used here for classification. I use LDA to predict the class of an item, in this case a word, with the help of numerical predictors, in this case, the corresponding word vectors. In this first study, I am always interested in the classification into two classes: for the bases, whether it is a base for an *-ity* or a *-ness* derivative, and for the derivatives, whether they are *-ity* or *-ness* derivatives. It is an instance of supervised classification, because the number of classes that the classifier is supposed to identify was decided on beforehand. To evaluate how well this classification works, the whole dataset is divided into a training set and a test set. The LDA classifier is trained on the training set, having pairs of vectors and their class available, and then evaluated against the test set, where the classifier predicts the class of the vectors.To assess the quality of the predictions, the results are checked against the true value in the dataset. The average weighted F1-score is used to quantify the success. The F1 score in its simple form (sometimes just called F-score), cf. (7), takes into account the precision and recall of the classifier.

(7) $$\text{F1} = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

Precision is the ratio of items that are correctly predicted for a given category, divided by the total number of items for which the classifier predicts this category. Recall is the ratio of the number of correct predictions of items in a category divided by the total number of items in this category. This is best explained with the help of an example: Let's say we have 250 word vectors representing 100 *-ity* and 150 *-ness* derivatives. We use our LDA classifier to predict for each word vector whether it belongs to a *-ness* derivative or not. We then compare the predictions to what we have in our dataset. This is usually done via a confusion matrix, cf. table 4. This matrix shows us how many word vectors were correctly predicted to correspond to *-ness* words, the true positives, and how many were wrongly predicted to be *-ness* words but are in fact not *-ness* words, the false positives. And we also have the false negatives, word vectors that are predicted to be not *-ness* but are in fact *-ness*, and the true negatives, word vectors that are correctly predicted to not be *-ness* derivatives.

Table 4: Toy confusion matrix for an LDA classifier predicting the class, *-ity* or *-ness*, for 250 vectors

Actual distribution		predicted -ness	predicted not -ness	total
-ness	150	100	50	150
-ity	100	30	70	100
column total	250	170	120	250

Precision quantifies the number of correct predictions and is the ratio of correct predictions to positive predictions overall, cf. (8) with the numbers from the table:

(8) $$\text{precision} = \frac{\text{true positives}}{\text{true positives}+\text{false positives}} = \frac{100}{100+30} = 0.77$$

Recall quantifies the number of correct predictions relative to the number of correct predictions that would have been possible. It is the ratio of true positives to the sum of true positive and false negatives, cf. (9).

(9) $$\text{Recall} = \frac{\text{true positives}}{\text{true positives}+\text{false negatives}} = \frac{100}{100+50} = 0.67$$

So, in our case the F1 value assessing the correct prediction of *-ness* derivatives is:

(10) $$\text{F1} = 2 * \frac{(precision*recall)}{(precision+recall)} = 0.71$$

Using the same toy data to calculate the F1 value for the correct prediction of *-ity* items, Precision is 70/120 = .58 and Recall 70/100= .7, yielding the F1 score .64. The weighted F1 score is the combination of these two scores in proportion to the number of occurrences of *-ity*/*-ness* in the dataset:

(11) $$\text{weighted F1} = 100/250 * \text{F1}_{\text{ity}} + 150/250 * \text{F1}_{\text{ness}} = 0.68$$

The average weighted F1-score reported below is based on a repeated stratified cross validation, using 10-splits and 3 repetitions. This means that the data is split into ten equal parts (or folds), and then the LDA classifier is trained on 9 parts and tested against the remaining 10th part. This is done 10 times, with each fold of the dataset being the test set once. This whole procedure is repeated three times, with each time a different split into ten parts.

This score is compared against the average weighted F1-score of a baseline classifier that either assigns the most frequent category to everything, or either of the two categories in the case of an exact split.[1] For our toy example above, the baseline classifier would assign *-ness* to every vector, because it is the most common category. With precision at

[1]For the unselected category, precision and recall were in both cases by default set to zero. This follows the reasoning on division by zero in `https://github.com/dice-group/gerbil/wiki/Precision,-Recall-and-F1-measure`.

150/250 = .6 and recall at 150/150 = 1, the F1 score is .75. For -ity, the F1 score is 0. The weighted F1 score is .45. That is, the classifier in our toy example clearly outperforms the baseline classifier.

All calculations of cosine similarities and vector manipulations were done with Python, the software implementations of both t-SNE as well as LDA I use come from Python's scikit-learn library (Pedregosa et al., 2011). All statistical analysis of cosine similarities was done with R Statistical Software (v4.1.2; R Core Team 2021), using the `mgcv` package for beta regression (Wood, 2017) and and `visreg` (Breheny and Burchett, 2017) for visualization.

3.4 Results [study 1]

Figure 2 shows the t-SNE visualization of the 300-dimensional vectors of the adjectival bases on a two-dimensional plane. Blue circles represent the projections of the vectors of the bases of *-ness* derivatives, red crosses represent the vectors of bases of *-ity* derivatives.

We see that each set of bases clearly clusters together, with *-ity* bases concentrated in the right half, *-ness* bases in the left half. There are bases of both types outside of their clusters and intruding into the other cluster, overall in equal measure for both *-ity* and *-ness* bases. The existence of these two clusters shows that the bases of *-ity* derivatives are clearly semantically distinct from the bases of *-ness* derivatives. The clear clustering is supported by the LDA classifier, trained to predict the vectors as either *-ity* or *-ness* bases: the average weighted F1 score is 0.849 (0.018 std), against the weighted F1 of 0.395 for a baseline classifier. In other words, the LDA performs well in classifying the bases into *-ity* and *-ness* bases, while the baseline classifier performs very poorly.

Figure 3 shows the t-SNE visualization of just the *-ive* bases within the set of non-doublets, with again blue circles representing projections of the vectors of *-ness* bases and red crosses projections of the *-ity* bases. In Figure 3, the *-ness* bases cluster in the upper half, the *-ity* bases in the lower half (note that the orientation of the clusters on the

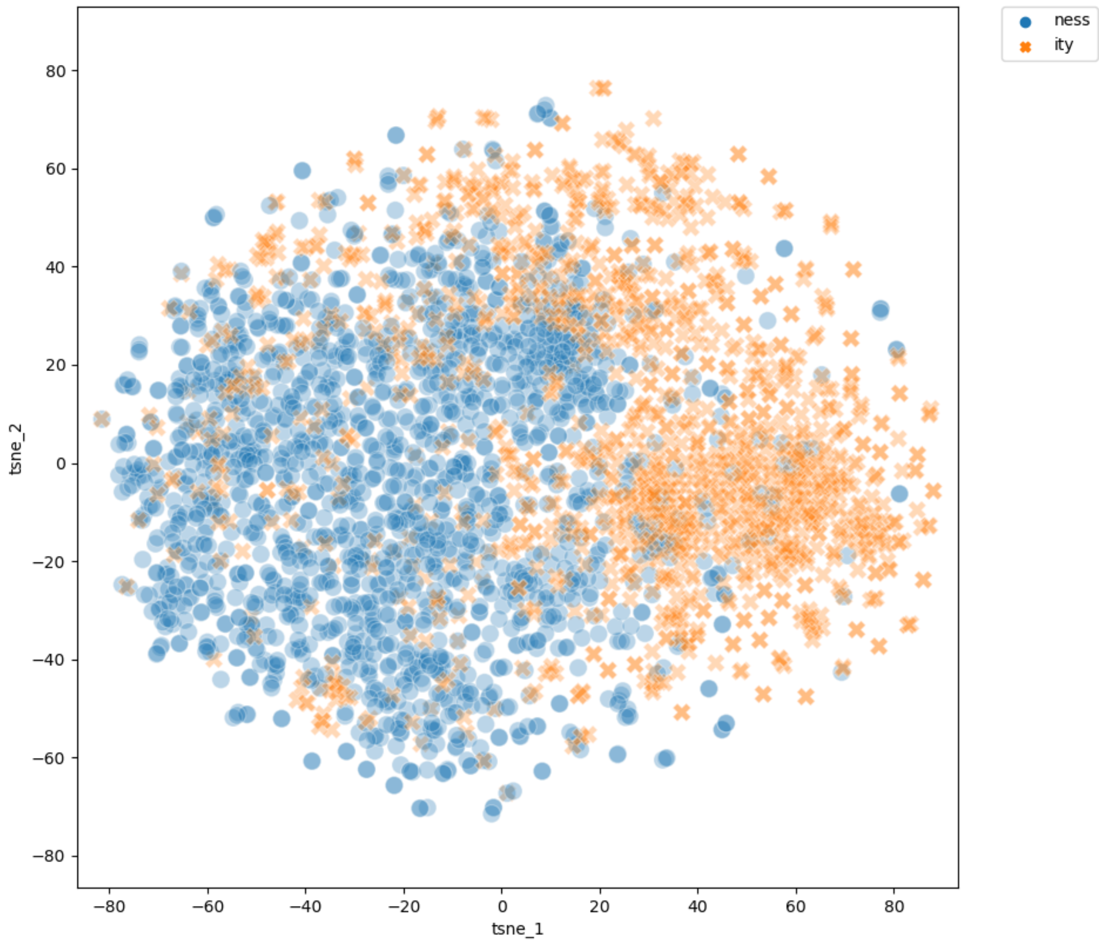

Figure 2: Projection of the vectors of the adjective bases into two dimensional space using the t-SNE dimension reduction technique. Bases of doublets are excluded.

plane is meaningless). We see that even if we restrict the dataset to bases ending on the same end-string, *-ive*, both types of bases are again clearly separated. So, even when no form-based feature allows for a distinction (all bases share the same form in that they end on *-ive*), there is a clear semantic difference between bases of *-ity* derivatives and bases of *-ness* derivatives. Again, the LDA similarly shows a high mean weighted F1 score of 0.744, standard deviation of 0.098, against the weighted F1 score of the baseline classifier of 0.385.

Turning now to the projections of the derivatives themselves, figure 4 shows the clustering of all derivatives, figure 5 that of only the derivatives with *-ive* bases. Except for different placements of the two clusters in the two-dimensional space, the overall pattern,

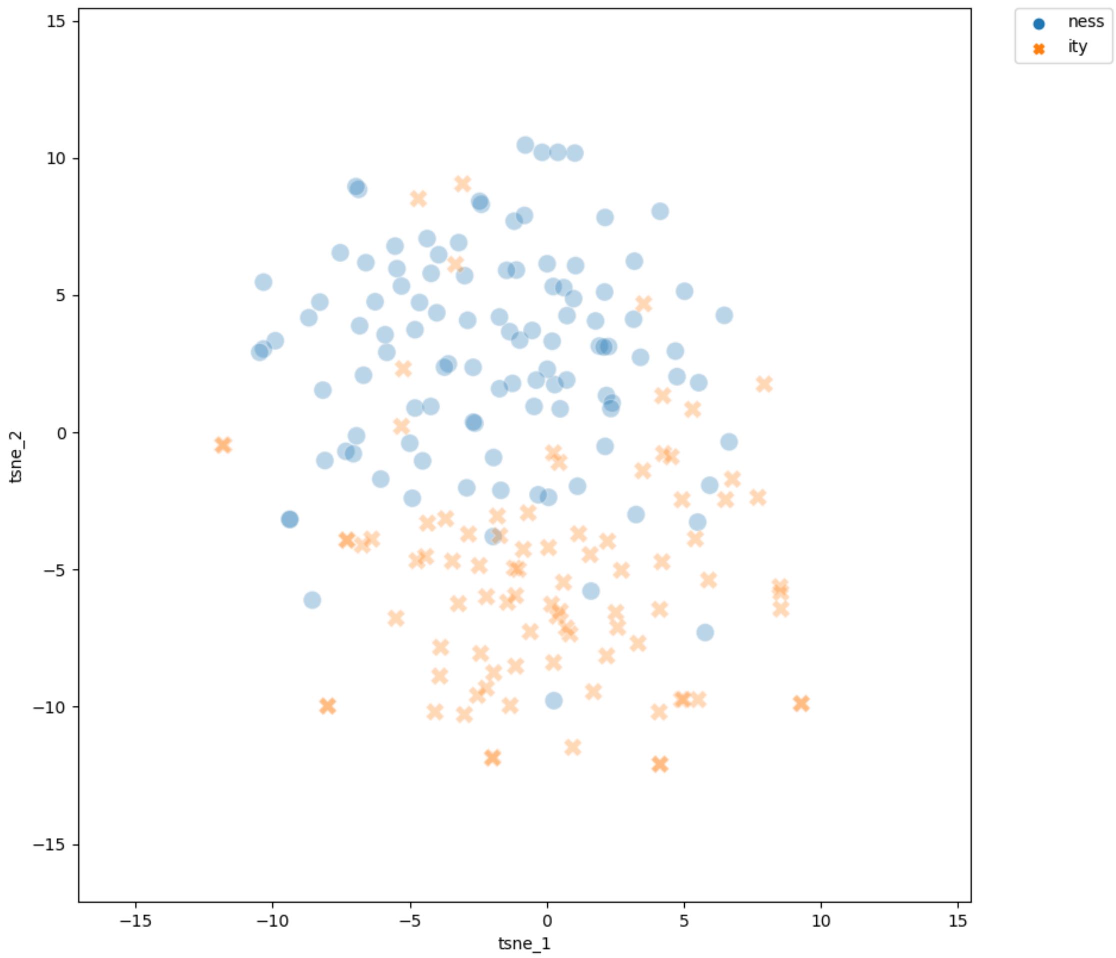

Figure 3: Projection of the vectors of the adjective bases ending on *-ive* into two-dimensional space using the t-SNE dimension reduction technique. Bases of doublets are excluded.

that is, two big clusters, is very similar to the pattern observed for the bases.

Again, LDA supports these clusterings, both for all derivatives (mean weighted F1 score = 0.858, std = 0.017), as well as for the *-ive* derivatives (mean weighted F1 = 0.863, std = 0.069).

Across frequency bands, the clustering into two main clusters of *-ity* and *-ness* vectors remains relatively stable, except for the two highest frequency bands, where *-ness* items are relatively rare (cf. the figures in the Appendix 1). Importantly, the clustering for base vectors and for derivative vectors is similar within each frequency band, as can also be seen in the LDA summary table in 5, which, for reference, repeats the value for the full dataset in the top row.

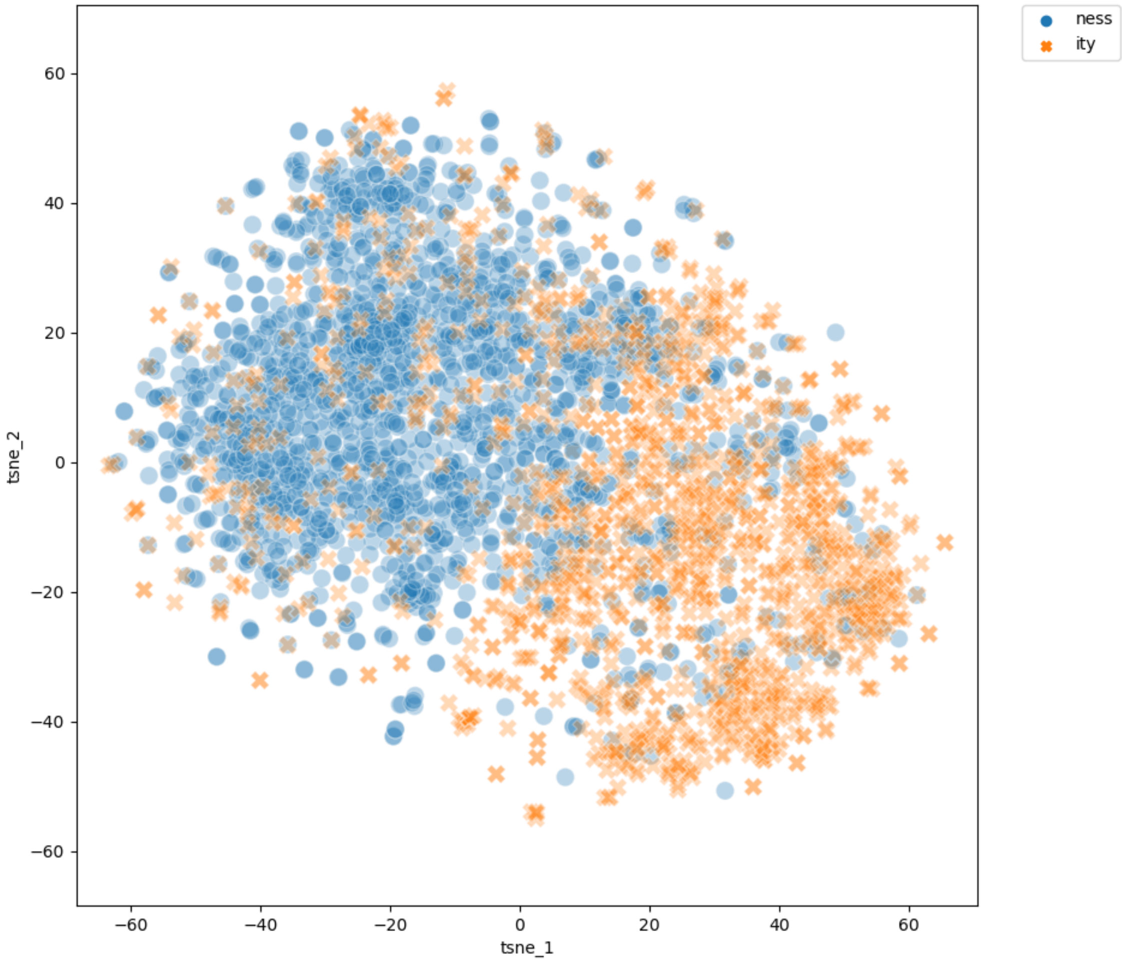

Figure 4: Projection of the vectors of the non-doublet derivatives into two dimensional space using the t-SNE dimension reduction technique. Derivatives forming doublets are excluded.

Table 5: Overview of LDAs across frequency bands for both bases and derivatives, reporting the mean weighted F1-score and the standard deviation. The weighted F1-score of the baseline classifier is given in the rightmost column.

	base vectors		derivative vectors		
subset	mean weighted F1	std	mean weighted F1	std	baseline
all	0.849	0.018	0.858	0.017	0.395
ultraLow	0.686	0.061	0.693	0.066	0.393
low	0.816	0.04	0.821	0.032	0.527
mid	0.753	0.057	0.747	0.057	0.426
midHigh	0.665	0.079	0.683	0.08	0.333
high	0.678	0.072	0.645	0.094	0.341
superHigh	0.756	0.075	0.777	0.102	0.584
ultraHigh	0.785	0.09	0.805	0.087	0.772

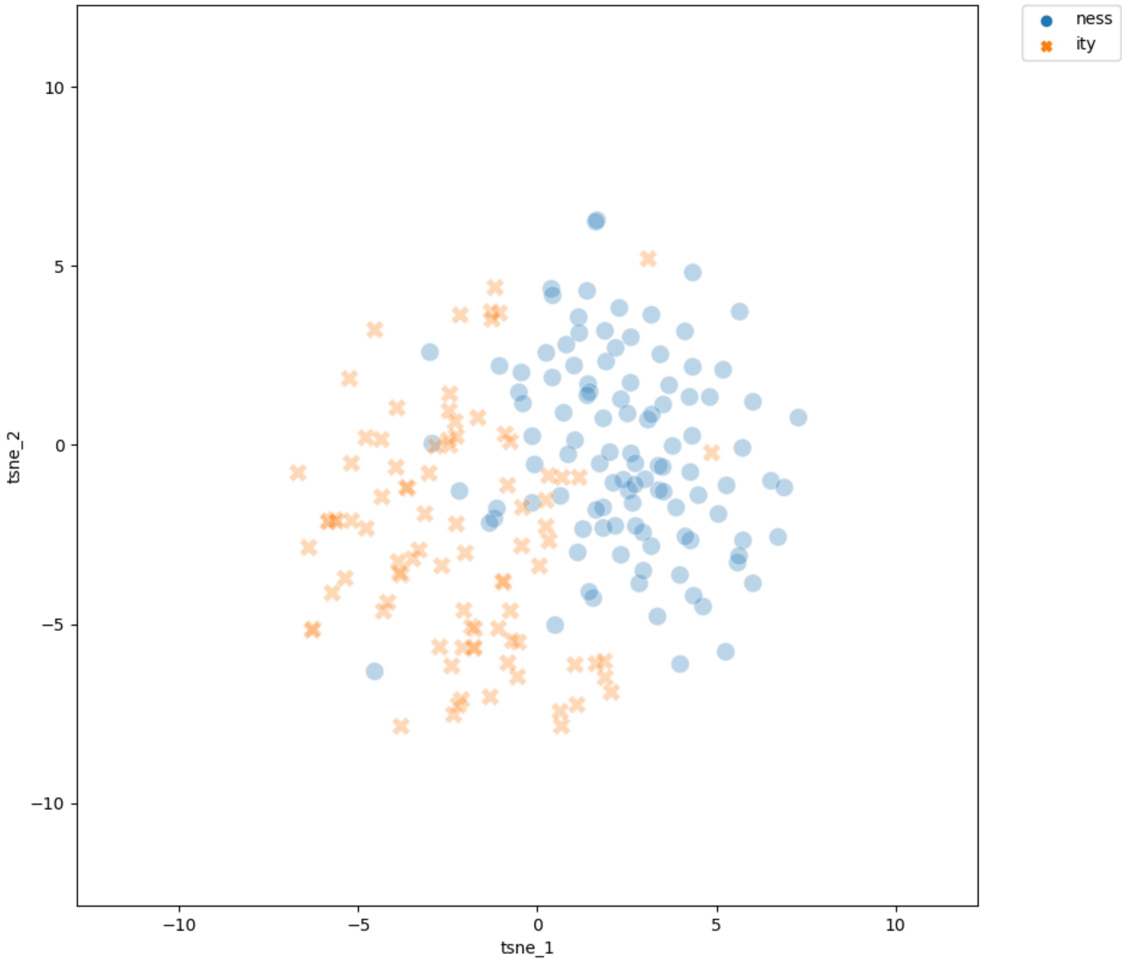

Figure 5: Projection of the vectors of derivatives of adjectives ending on *-ive* into two dimensional space using the t-SNE dimension reduction technique. Derivatives forming doublets are excluded.

LDA performs best on the whole dataset, and across frequency bands the mean weighted F1 score never drops below 0.665 (median = 0.753) for the base vectors and 0.645 (median = 0.747) for the derivatives vectors, and the values are highly correlated (Kendall's tau = 0.905, p = 0.003). In contrast, the baseline classifier performs poorly across the board, with the exception of the superHigh and ultraHigh frequency bands, where the scoring profits from the high imbalance of *ity* to *ness* items (superhigh: 115 to 48, ultrahigh 97 to 18, cf. Table 2). However, even there the LDA scores are better.

When we compare the cosine similarities between the *-ity* and *-ness* base-derivative pairs, we observe considerable variation for both the *-ity* and *-ness* pairs. Descriptively, they are globally similar, cf. the values characterizing the distribution in table 6, further illustrated in the density plot in figure 6.

Table 6: Cosine similarities between the base and the derivative for all *-ity* and *-ness* pairs (excluding doublets)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-ity-pairs	0.1504	0.5034	0.5908	0.5764	0.6572	0.8883
-ness-pairs	0.1673	0.5164	0.5798	0.5659	0.6357	0.8624

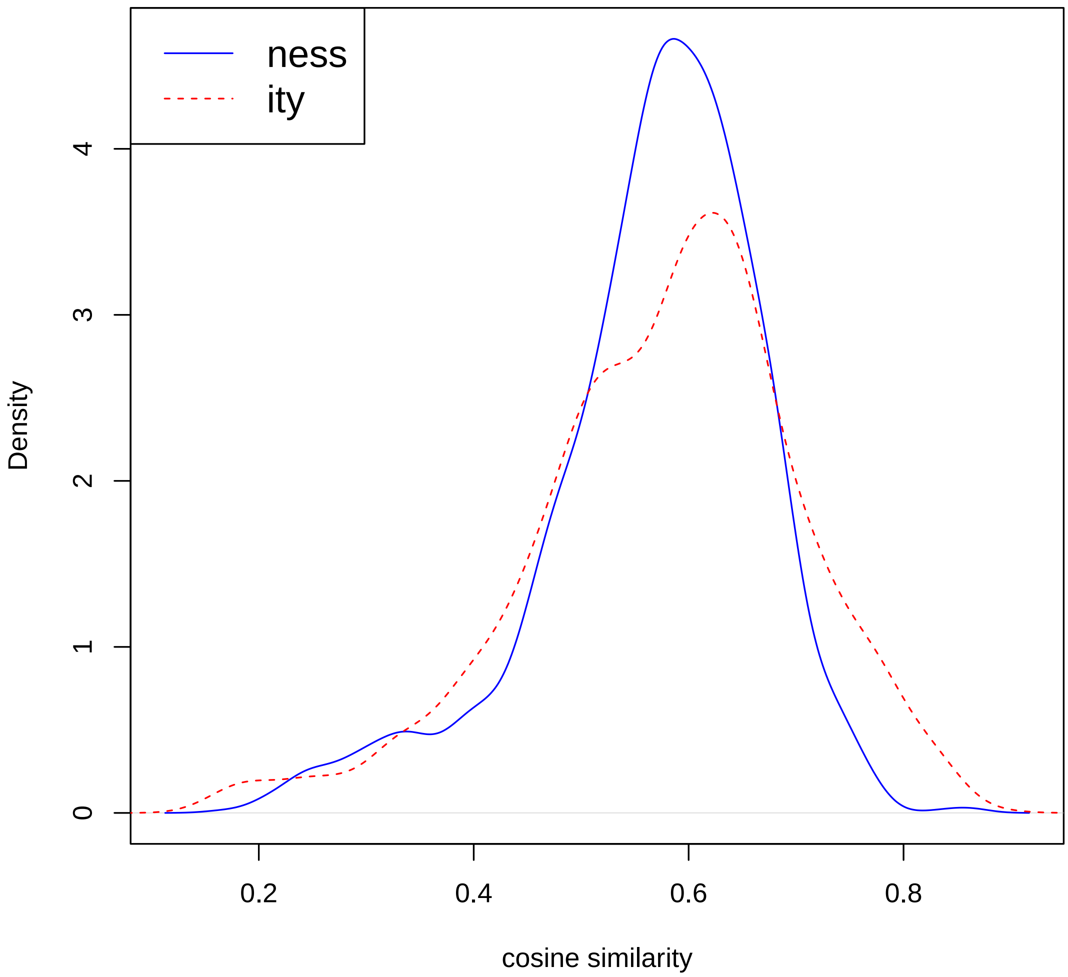

Figure 6: Density plots for cosine similarities between paired *-ity* and *-ness* bases. Derivatives forming doublets are excluded.

Both are not normally distributed and show a similar range of variation. Even so, they significantly differ statistically both with regard to their mean as well as their variance

(Wilcoxon: W = 1189258, p-value = 0.004666; F-test: F = 1.5035, p-value = 2.887e-15).

When taking into account the role of lemma frequency here with the help of a beta regression model, cf. table 7, we find that the weakly correlated log-frequencies (Kendall's tau = 0.301, p = <2.2e-16) of bases and derivatives emerge as highly significant predictors and participate in an interaction. Lemma frequency accounts for almost all of the variance explained by the model. The affix itself is also a significant factor, but contributes very little to explained variance (its inclusion leads to an improvement of the adjusted R-squared by just 0.0016).

Table 7: Beta regression for cosine similarity between the non-doublets. R-sq.(adj) = 0.09 Deviance explained = 9.81%

Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.961831	0.045961	20.927	<2e-16
affixness	0.036753	0.017597	2.089	0.0367
derLogFreq	-0.085729	0.009922	-8.640	<2e-16
baseLogFreq	-0.110019	0.006458	-17.035	<2e-16
derLogFreq:baseLogFreq	0.013741	0.001172	11.723	<2e-16

In other words, there is a very slight influence of the affix as such on the observed cosine similarity.

3.5 Discussion [Study 1]

The patterning of the data in Figure 2 shows that the bases of *-ity* and *-ness* derivatives fall into two clear clusters. This is in line with hypothesis (1a) and shows that the meaning of the base is a very powerful predictor of whether the base selects for *-ity* or *-ness*. Addressing hypothesis (1b), Figure 3 shows that this even holds for the *-ive* bases, a test set where the last syllable does not contain any useful information for morphological or form-based approaches. The results for *-ive* bases very clearly show that the observed effects are independent of the morphological and formal makeup of the bases. This result is in line with the points made by Riddle (1985) with regard to small semantically consistent swathes of the lexicon showing clear preferences (e.g. the color words) and the idea that the morphological endings of bases might be useful as formal predictors because they serve as indicators of shared semantics. In fact, the results indicate that the role of base semantics extends further: a difference in base semantics is actually the pervasive feature that holds for most of the bases, even trumping the influence of shared morphology and/or form as in the case of *-ive*.

Turning to the derivatives, figures 4 and 5 show that the derivatives also fall into two main clusters, in this respect similar to their bases. When looking at the cosine similarities between bases and derivatives, the overall distributions are globally similar in that they both are not normally distributed, both left-skewed and with very similar medians and means. But they are significantly different, and when modeling the distribution, although the interaction of base and derivative frequencies accounts for almost all of the explained variation, the affix still emerges as a significant factor. How does this link to hypothesis (2a), to what extent can we say that the change in meaning induced by the two affixes is the same? It is obviously not exactly the same, but I believe that overall we can say that the change in meaning they adduce is still relatively similar. The affix emerges as significant predictor, but it accounts only for a minimal part of the variation. At the same time, saying that it is relatively similar does not mean that we can identify a very specific change in meaning that is induced, since part of what makes the two similar is the huge variation and also the skewedness of the distributions. And what does it mean for the meaning change induced by the affixes that both bases and derivatives show comparable distinct clusterings? Together with the cosine similarities, it can again be taken to support the view that the meaning change is similar: the effect of adding either affix does not make the resulting derivatives more or less similar than their respective bases already are, that is, they are moved in the distributional space, but these shifts are similar for both affixes.

This interpretation is not in line with regard to the hypothesis from Baeskow (2012) that the derivatives also show a ‘syntactic effect’ in that *-ness* derivatives favor the syntactic realization of the external argument of the adjective base. This would lead us to expect a clearer difference between the vectors of the derivatives than between the vectors of the bases, and also a clearer pattern in the cosine similarities, something we do not find. In a similar vein, this finding is also unexpected on the assumption of a pervasive register difference between *-ity* and *-ness* derivatives: again, we would expect a clearer difference between the vector of the derivatives, and also a clearer pattern in the cosine similarities.

4 Study 2: Doublets

Study 2 considers all (and only) doublets. Doublets are pairs of *-ity* and *-ness* derivatives formed from the same base. The hypothesis in focus here is (2b): If *-ity* and *-ness* are synonymous, doublets (such as *aggressivity/aggressiveness*) should show no systematic meaning difference, as they share the same bases. Here, I again use the same techniques as before for the analysis of the derivatives: First, t-SNE for visualization and LDA for statistical corroboration of the derivative vectors. Second, I also consider the similarities between the doublets in order to see whether any general trend can be identified in the data. Derivative frequencies are also considered. If frequency is correlated with lexicalization, then I expect a higher likelihood of higher frequency items having developed idiosyncratic meanings.

4.1 Study 2: Materials and techniques

4.1.1 Materials

There are 130 doublets in the data. Just as for the non-doublets, the doublets show clear differences between the *-ity* and *-ness* items in their distribution across the frequency spectrum, as shown in the density plot in Figure 7.

Just as for the non-doublets, the *-ity* derivatives tend to have higher token frequencies, whereas the *-ness* derivatives on average have lower frequencies and only very few high frequency items. The frequencies of *-ity* and *-ness* items within the doublets are not correlated (Kendall's tau = -0.02, p = .68).

When we look at the endings of the bases, cf. Table 8, we expectedly see a smaller

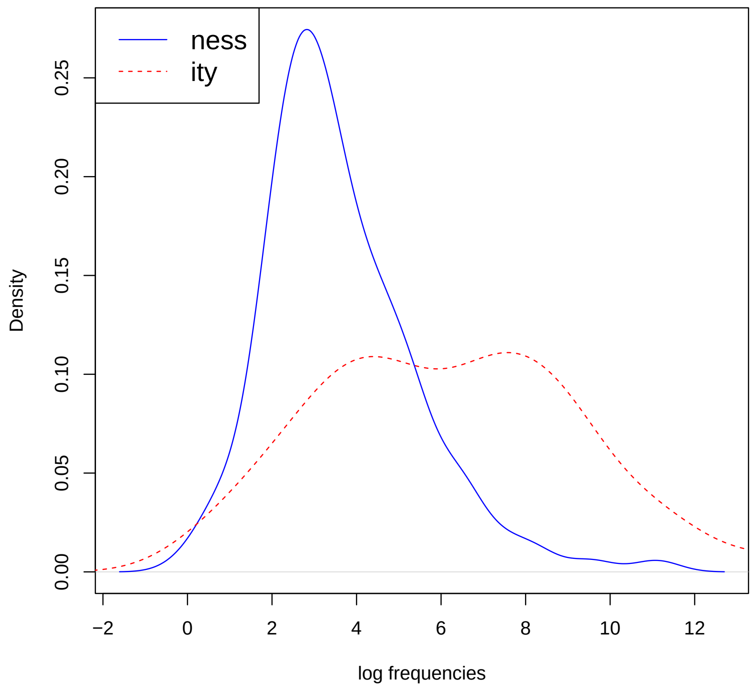

Figure 7: Density plot of the log-frequencies of the *ity* and *ness* derivatives in the doublets.

subset of possible adjective endings, with none of the 100% *-ness* endings from study 1 occurring here. Notably, 47 doublets are derived from adjectives ending in *-ive*.

Table 8: Distribution of endings in the bases of all doublets

base ending	number of doublets	example
al	13	*factuality/factualness*
ar	1	*clearity/clearness*
ble	9	*comfortability/comfortableness*
ic	4	*genericity/genericness*
id	6	*morbidity/morbidness*
ive	47	*productivity/productiveness*
ous	24	*audacity/audaciousness*
te	1	*chastity/chasteness*
OTHER	25	*density/denseness*

4.1.2 Techniques

For looking at the vectors themselves, I use the same techniques as in study 1, that is, t-SNE visualization combined with LDA. Further, to explore the within-doublet similarities, I look at the cosine similarities between members of a doublet. In order to probe the role of lexicalization via lemma frequency, I use beta regression to model the similarity within doublets.

4.2 Results [study 2]

Figure 8 shows the projection of all derivatives participating in doublets, again using the t-SNE visualization. In sharp contrast to the projections in study 1, there is no discernable clustering of the *-ity* and *-ness* vectors. The LDA results in a notable low mean weighted

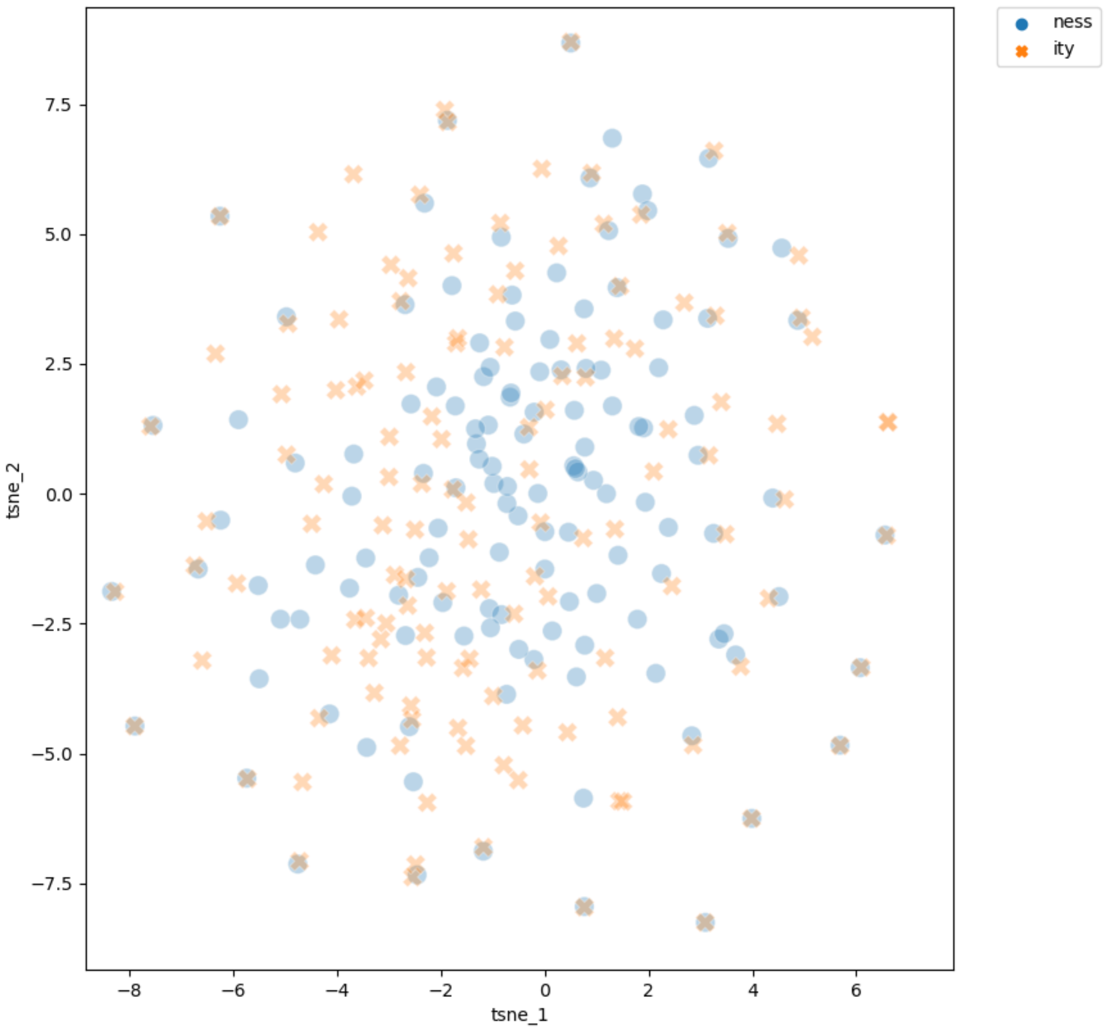

Figure 8: Projection of the vectors of all doublets into two dimensional space using the t-SNE dimension reduction technique.

F1 score of just 0.562 (0.101 std), against a baseline classifier weighted F1 score of 0.333.

When it comes to the cosine similarities within doublets, we observe considerable variation, with a minimum similarity of 0.268 and a maximum of 0.867 (median = 0.639, mean = 0.614). Table 9 illustrates the data by showing doublets across the distribution of cosine similarity values.

Table 9: Illustration of doublets across the distribution of cosine similarities within doublets. The two doublets closest to the respective cosine similarity values have been selected.

place within the distribution of cosine similarities	doublet
Min (0.2680)	*opportunity/opportuneness* *casuality/casualness*
1st Qu. (0.5416)	*naturality/naturalness* *obliquity/obliqueness*
Mean (0.6129)	*chastity/chasteness* *changeability/changeableness*
3rd (0.7212)	*exhaustivity/exhaustiveness* *passivity/passiveness*
Max (0.8671)	*impassivity/impassiveness* *inclusivity/inclusiveness*

This variation is not caused by any specific pattern regarding the end-string. When just considering the largest subset with the same ending, again *-ive*, we find a similarly-shaped wide distribution across the cosine similarity space, cf. Table 10.

Table 10: Distribution of cosine similarities within *-ive* doublets

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.3252	0.5578	0.6575	0.6357	0.7265	0.8671

When exploring the role of lemma frequency here, we find that the uncorrelated log-frequencies of *-ity* and *-ness* items within the doublets show an interaction, cf. the beta regression model in 11 and the interaction plot in Figure 9.

The interaction plot shows in its first three panels that for very low up to medium high *-ness* frequencies, higher *-ity* frequencies are associated with lower cosine similarity values of the doublets. These negative associations all come with tight confidence bands, showing that this negative association is reliable. Doublets that illustrate this effect are *officiality/officialness* vs. *morbidity/morbidness* with cosine similarities of 0.82

Table 11: Beta regression for cosine similarity between the doublets. R-sq.(adj) = 0.137 Deviance explained = 16.4%

Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.400541	0.236732	5.916	3.30e-09
ityLogFreq	-0.168542	0.038275	-4.403	1.07e-05
nessLogFreq	-0.164697	0.055822	-2.950	0.00317
ityLogFreq:nessLogFreq	0.031672	0.009653	3.281	0.00103

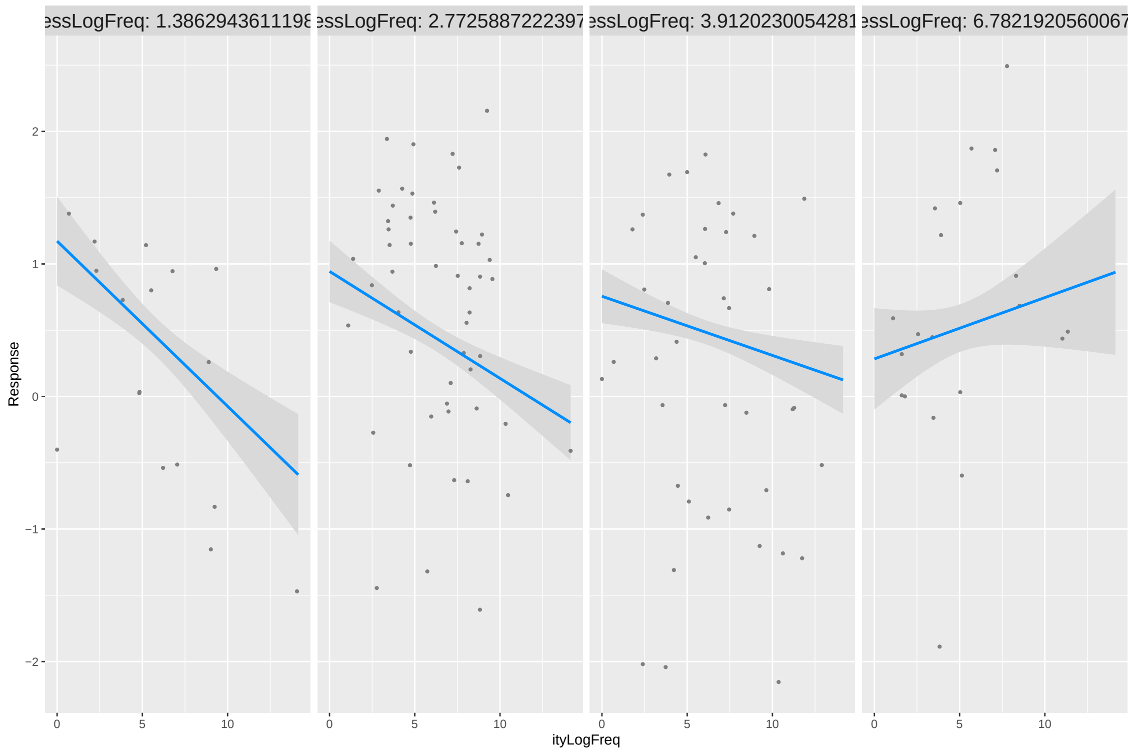

Figure 9: Interaction plot of log frequencies of *-ity* and *-ness* derivatives in the beta regression model for doublet similarity. The individual panels show the relationship between the *-ity* log-frequency and the cosine similarity within doublets for specific values of *-ness* log frequencies. These *-ness* log frequencies values increase across the panels, from left to right, as shown at the top of the panels.

and 0.41. Both *officiality* and *officialness* come from the ultraLow frequency band. In contrast, *morbidity* is from the ultraHigh frequency band, while *morbidness* again is an ultraLow frequency item. Across the panels, the slope of the negative association and its starting point change: the highest cosine similarity predicted by the model becomes successively lower, and the slope successively less steep. However, this second aspect is less reliable: If we concentrate on the effect of *-ness* frequency when *-ity* frequency is low, that is, on the leftmost points in each of the three panels, we see that each starting point is contained in the other two panels' confidence intervals. Similarly, at the end point on the right hand side, the confidence intervals overlap from panel to panel. For the high frequency end of the *-ness* items, shown in the rightmost panel, the relationship changes and a positive association is plotted. Here, the model is very unsure about the association, as shown by the huge confidence intervals. Pairs such as *inclusivity/inclusiveness* and *reasonability/reasonableness*, with cosine similarities of 0.87 and 0.73, illustrate this panel, with *inclusivity* in the superhigh bracket and *inclusiveness* in the high bracket, but *reasonability* in the low bracket and *reasonableness* again in the high bracket.

4.3 Discussion [study 2]

Study 2 lets us address hypothesis (2b): Figure 8 reveals no clustering, again in line with the idea that *-ity* and *-ness* are synonymous. At the same time, the range of cosine similarities within doublets is in line with the observations from the literature that there are doublets which clearly show a meaning difference in specific contexts or for which one would even be hard pressed to find a context where they could be used interchangeably. But, and this backs the points by Bauer, Lieber, and Plag (2013), there is no systematic pattern across the doublets. The similarity within doublets can be successfully modeled with the help of the frequencies of the participating *-ity* and *-ness* forms. This interaction is difficult to interpret, and I will restrict the interpretation here to the one single clear effect: the role of *-ity* frequency when *-ness* frequency is low to midhigh: cosine similarity is negatively correlated with *-ity* frequency. This is in line with the idea expressed in Bauer, Lieber, and Plag (2013) that high frequency *-ity* items are associated with a large number of different lexicalization pathways, moving the corresponding vectors away from more conventional usages. That is, if we assume that high-frequency items are in general liable to become lexicalized, and if we further assume that there is not only one possible lexicalization pathway for *-ity* derivatives, then we would expect (a) that high frequency items are used differently from low frequency items and (b) that these different usages show not one, clearly discernable, pattern, but only more or less idiosyncratic differences. If we complementarily also assume that the same is to be expected for *-ness*, that is, the higher the frequency, the more often we find different lexicalization pathways, we would expect that for low frequencies, both *-ness* and *ity* derivations show a more conventional picture. The high cosine similarity here is then again supporting the idea that the two suffixes have the same semantic effect on their bases. That the effect of *-ness* frequency given a low *-ity* frequency is somewhat less strong in comparison could indicate that *-ness* derivatives show less variety in their lexicalization pathways. All in all, the reliable patterns in the results for the doublets are in line with the idea that *-ity* and *-ness* induce similar meaning shifts and that diverging usages of doublets are likely to have arisen from lexicalization.

5 General discussion

5.1 Prediction via bases and the relationship of semantic vectors to other predictors

One core result of this study is that the distributional vectors of the bases already predict whether the base comes with a corresponding *-ity* or *-ness* derivative. What exactly are then then the semantic properties shared by the respective bases, and how does this relate to morphological and form-based prediction?

Mapping the data onto traditional semantic classes turns out to be problematic. This is due to the fact that there are no off-the-shelf classifications of adjectives that allow meaningful distinctions across the 3,277 adjectival bases considered here. This holds for classifications used in theoretical works as well as for classifications used in natural language processing. Intuitively clear classes, for example the speed class introduced in Dixon (1982) or the class of color words discussed by Riddle (1985) yield extremely small subsets, while most adjectives, in terms of Dixon's classification, can be assigned to the human propensity class, which thus is not diagnostically helpful. And a large lexical database like WordNet (Fellbaum, 1998) has only a crude distinction into two adjective classes, relational vs. all other adjectives. As a result, studies using distributional semantics to explore adjective semantics often look at a very small number of adjectives, cf. Schäfer (2023) who uses the Dixon-classification and investigates 11 human propensity and 11 speed adjectives. At least we can use these small sets as a proof of concept that the t-SNE algorithm does cluster semantic features. In Figure 10, this is shown for the adjectives from the two classes explored in Schäfer (2023) that are also in my dataset, along with all color adjectives in the dataset. While admittedly these 24 bases are an extremely

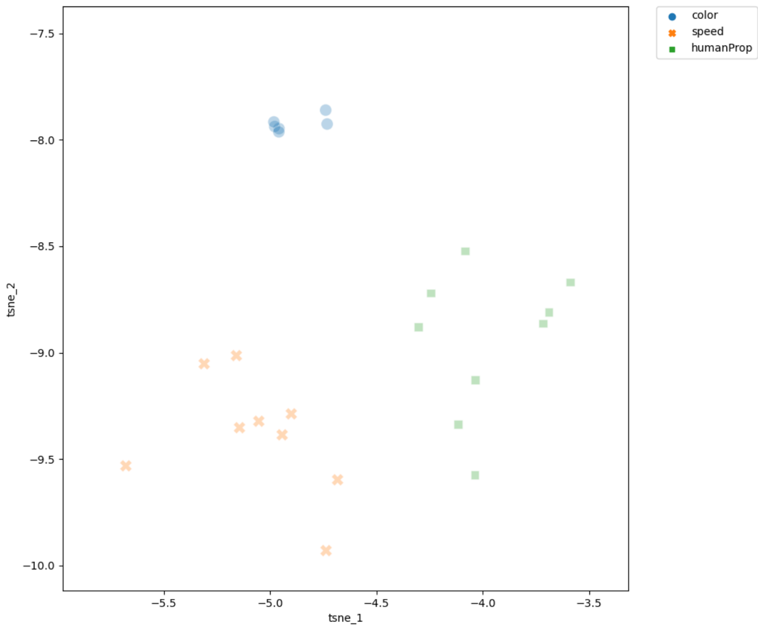

Figure 10: Projection of adjective bases from three different semantic classes of adjectives: color (6 bases), human propensity (9 bases), and speed (9 bases) into two dimensional space using the t-SNE dimension reduction technique.

small subset of our dataset, they show that traditional semantic classes are in fact recognized by the t-SNE algorithm. For my *-ity*/*-ness* dataset, we can therefore assume that these traditional classifications are part of the structure that is exploited in the vectorspace for clustering the vectors and for the LDA-predictions.

Moving away from the traditional semantic classes, there is the question of how the structure in the distributional vectors exploited here relates to the classic findings with regard to the influence of the morpheme-makeup of the bases and/or their form. Starting with the morphemic endings associated with either *-ity* or *-ness* derivatives, there is one obvious way in which to explore the semantic space further: we have Riddle's (1985) suggestion that the suffixes within the bases might actually influence the choice between *-ity* and *-ness* because the suffixes still have semantic significance. If this is the case, then we would expect that complex bases with affixes that are known to play a role in the selection, cf. the discussion in section 2.1 and also the overview by ending in table 3 for the non-doublets and in table 8 for the doublets, also cluster together within the distributional vectors of all bases. I will explore this here by looking at three endings that usually corresponds to morphemes, the big group of *-ble* bases, containing all bases that are formed with either *-able* or *-ible* as the group that most favors *-ity*, and two smaller groups of bases ending in *-ful* and *-ish*, containing all bases formed with the respective affixes and showing a clear preference for *-ness*.

Figure 11 shows the t-SNE visualization for *-ble* against all other base vectors in the dataset. In addition, the *-ble* bases that participate in doublets and those that go only with *-ness* are marked. We see that the only-*ity* bases form a large cluster towards the bottom of the plot, and several smaller ones in the upper half. Most clusters, including the biggest cluster, overlap with non-*ble* bases. The doublet and the only-*ness* bases do not form clusters and, in the majority, are relatively removed from the central cluster. That is, the majority of *-ble* bases that are only associated with *-ity* in my dataset are also close to each other in our vectorspace, while the more peripheral members, the doublet and *-ness* bases, do not form a consistent group and are removed from the core group of *-ble* bases. Overall, the finding that quite a number of only *-ity -ble* bases occur all over the vectorspace indicate that as a whole it is not forming a consistent group, by itself a likely explanation for the development of doublets and items with even an *-ness* preference.

For *-ness*, *-ful* and *-ish* are both only associated with *-ness* forms, there are no doublets nor instances of bases prefering *-ity*. Figure 12 shows that this is reflected very clearly in the corresponding semantic vectors: Both groups mainly cluster in the upper left third, and overall distribute over just a bit more than half of the vectorspace. Just as with *-ble*,

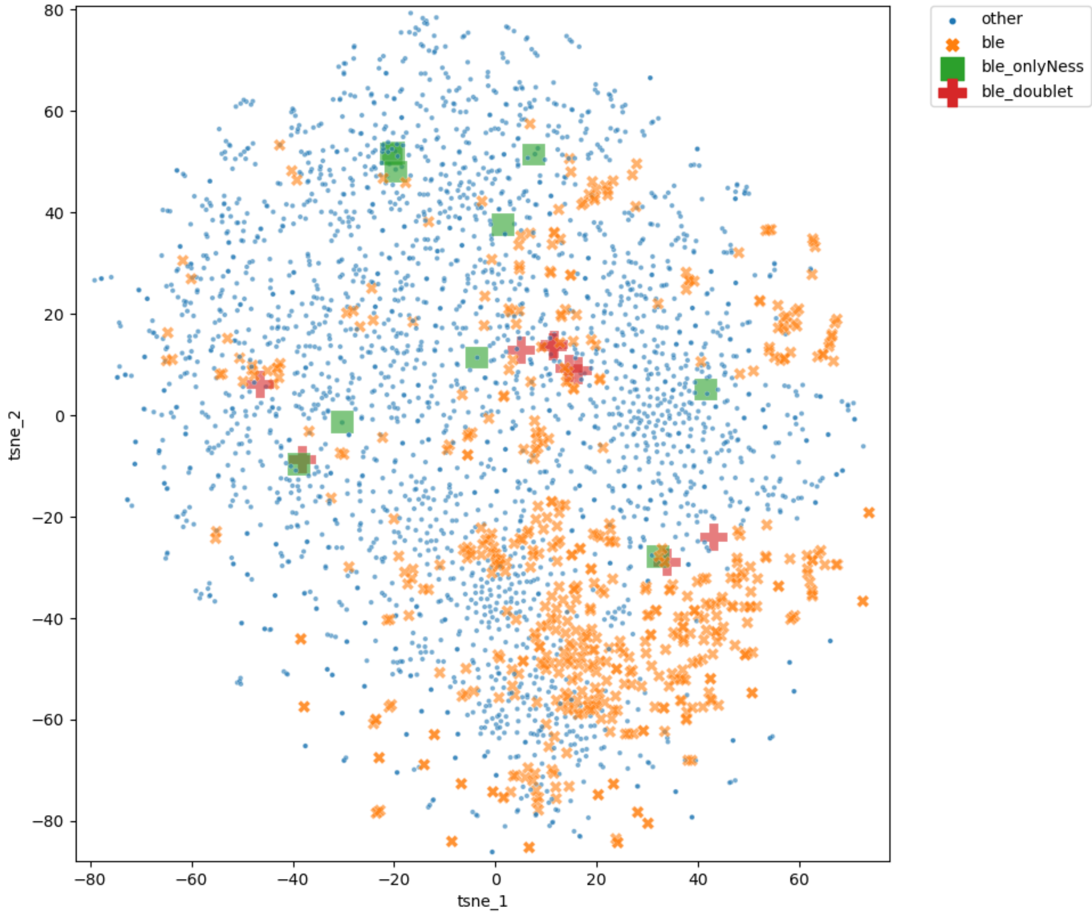

Figure 11: Projection of the vectors of all bases into two dimensional space using the t-SNE dimension reduction technique. All *-ble* bases are labeled

they do not occupy unique spaces within the bases, but overlap with each other and other bases. Even though both groups overlap, their overall clustering in such a clearly defined area of the vectorspace partially explains why the t-SNE visualizations and LDA predictions without the knowledge of the internal make-up of the bases already work so well: semantically, too, they are clearly set apart from roughly half of the other bases.

What about *-ive*? We have already seen the plots and LDA analysis of just its non-doublet bases. Figure 13 shows all *-ive bases* in the context of all other bases, again with the three groups, *-ive* bases that go with *-ity*, *-ive* bases that go with *-ness*, and doublet bases marked. If we look at the places of just the *-ive* bases in my dataset, we clearly see that they do not form any sort of clear cluster in the vectorset of all bases but are distributed with one relatively tight cluster of *-ity* bases in the left upper corner, and most *-ness* bases

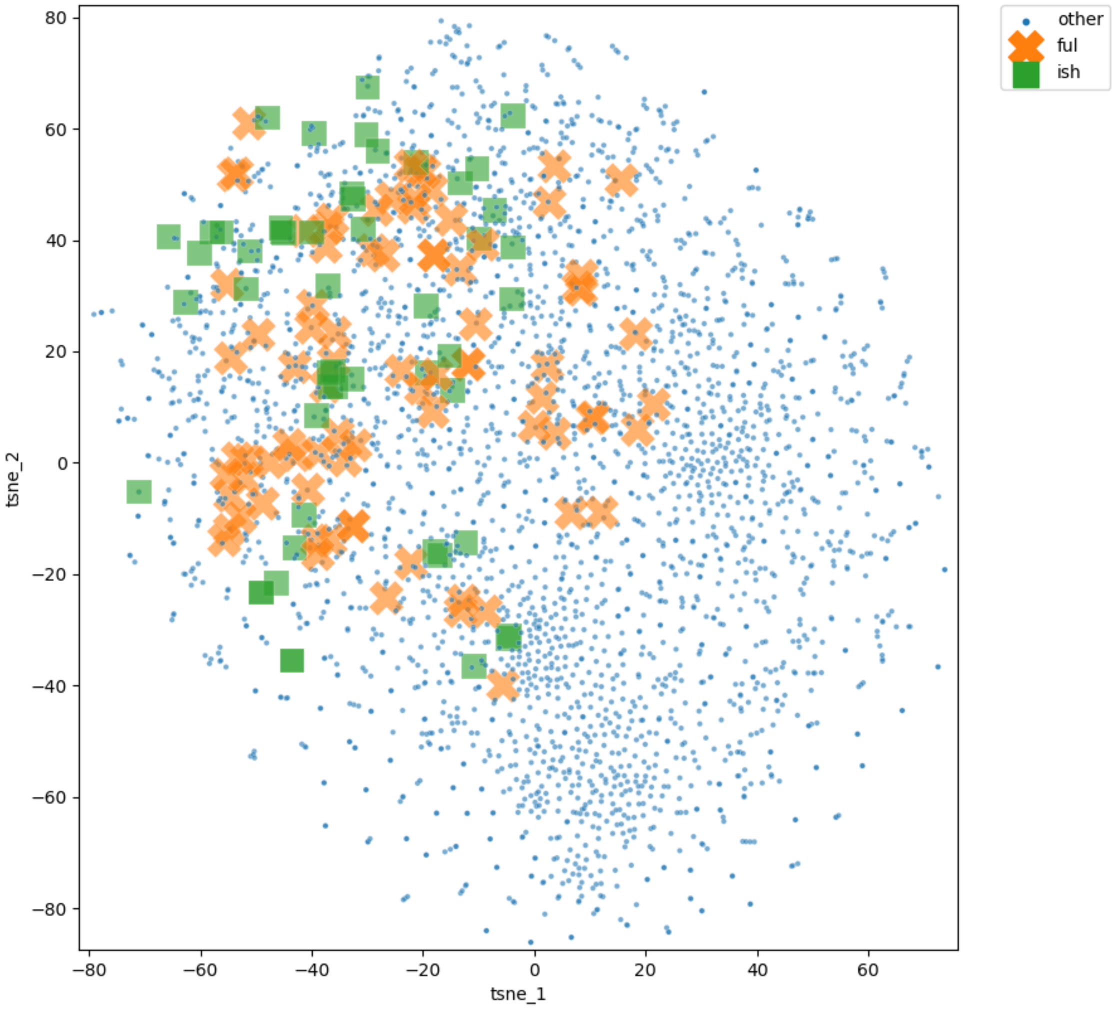

Figure 12: Projection of the vectors of all bases into two dimensional space using the t-SNE dimension reduction technique. All *-ful* and *-ish* bases are labeled

falling outside of the third of the graph extending from this cluster. While the picture for the doublet-bases is not absolutely clear, many fall in the middle area between these two regions or are located at the edge of the *-ness* area. So in terms of the distribution of its vectors, the reason why *-ive* is not a good predictor of an element taking either *-ness* or *-ish* is simply that within the semantic space of all base it does not occupy a clear cluster that would go with either of the two affixes.

Given these figures, it seems plausible to assume that in many cases including information about the endings would not yield better predictions, but this is an issue left to further research here: I have not looked at the placement of all affixes in the model, and I

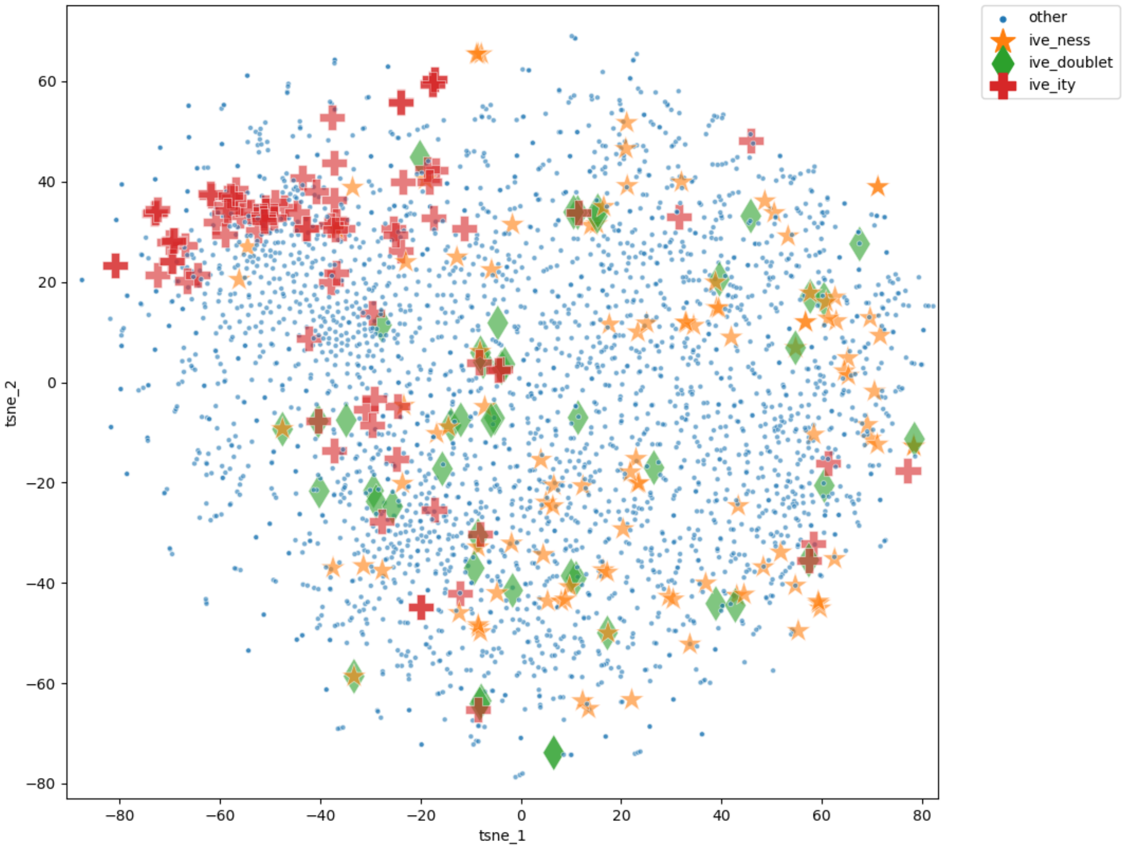

Figure 13: Projection of the vectors of all bases into two dimensional space using the t-SNE dimension reduction technique. All *-ive* bases are labeled

have also not distinguished between endings in terms of form and in terms of morphemes. The mean weighted F1 score for the LDA of the bases in the whole dataset without doublets is 0.849. This is slightly lower than the weighted F1 score of 0.887 for Arndt-Lappe's AM model trained and targeting OED 20th century neologisms, but our two studies cannot be compared directly. First, the set of OED 20th century neologisms is smaller than the dataset investigated here, with 344 *-ity* and 220 *-ness* derivatives in her dataset against the 1475 *-ity* pairs and 1802 *-ness* in my non-doublet dataset (note: Arndt-Lappe also models the OED neologisms of the two previous centuries, but not all taken together, and her model is best at modeling the 20th century neologisms). Second, the bases compared in my study are more tightly constrained in being all adjective bases, while Arndt-Lappe (2014) used a dataset with not only bases of different wordclasses but also non-word bases. The difference in word-status, i.e., word vs. non-word (e.g. phrase/bound form), was one of the predictors in the analogical model, the other predictors were purely form-based coding, in particular, six phonetic features describing the two final syllables of the base. It is an intriguing question how the two systems, my LDA based on distributional vectors and Arndt-Lappe's model using word-status and form-features, would perform on the respective other datasets. Given that the predictive power of Arndt-Lappe's model diminishes when she uses the neologisms of previous centuries (19th and 18th) to predict the neologisms of the following century, we can hypothesize that the base vectors might also form tighter sets when considered in comparable time brackets corresponding to first occurrence of the respective derivatives. A further open question is to what extent Arndt-Lappe's models were successfully able to deal with subsets with the same ending, in our dataset exemplified by the *-ive* non-doublets, in which the distributional vectors of the bases are still good predictors. Ideal, but left here for further research, would be a direct comparison of the different approaches on the same dataset, or even, as a first step, the investigation of the data used in my study with the help of vectors that are sensitive to the internal structure of the words, for example fastText vectors with subword information.

5.2 The issue of affix synonymy

The results with regard to affix synonymy have been less clear than the results concerning the role of base semantics in affix selection. In fact, as one reviewer points out, if we also approach synonymy from a purely distributional perspective, we would have to say that they are clearly not synonymous: they typically occur in different contexts, with the subset of doublet bases only constituting 4% of the full set of bases. Thus, a more moderate question to ask is whether they have the same effect on the meaning of the base. Here, study 1 showed that the semantic similarity in terms of cosine similarities between base and derivatives is relatively similar but nevertheless significantly different. We also saw from study 1 that the semantic effect of the affixes does not result in the sets of derivatives becoming more or less similar than the bases already were. From study 2 we saw very clearly that there is no consistent pattern in the doublets, that is, it does not look as if accounts that assume one function for *-ity* and one for *-ness* in the case of doublets could explain the distribution. All the same, given the huge variance in the vector similarities observed for both the set of non-doublets and the set of doublets, the assumption of one single meaning change consistently induced by either affix does not seem justified. A more detailed look at the offset vectors of the bases and derivatives, that is, the vectors that result from subtracting the vector of the base from the vector of the derivative, could yield more insights but is left to future studies here. Finally, I argued in the discussion of study 1 that the findings are not in line with a syntactic based explanation of the difference between *-ity* and *-ness* derivatives, nor with a register-based explanation: in both cases we would expect a clearer difference between the semantic vectors of the derivatives than what we find for the bases: if the contexts are more distinct for the derivatives, this should allow to separate the vectors better, and we would also expect a very clear difference in the cosine similarities between the two groups. Can I exclude an influence of register altogether? No, technically, for example, it could be that the semantic effect of adding either *-ity* or *-ness* in terms of traditional semantics, e.g., non-distributional semantics, makes the set of derivatives more similar. But if it is in fact true that *-ity* and *-ness* are associated with clearly different registers, this traditional semantic feature might be lost in the vector representation of the derivatives since they incorporate register information. In other words, maybe adding either affix makes the derivatives more similar than their bases when attempting a semantic description of their meaning that leaves register aside. But since distributional vectors as used here do not leave register information aside, such an effect could not be detected in my setup. Implausible as this seems to me, further research is needed to exclude this possibility.

6 Summary

This paper addressed the two core questions concerning the affix rivalry between *-ity* and *-ness* with the help of distributional semantics. Study 1 has shown that already the distributional vectors of the adjectival bases of *-ity* and *-ness* derivatives are good predictors of the affix choice. This does not only hold across all forms, but also for the only substantial subset of bases sharing the same ending, the *-ive* bases. Even for those, the bases associated with either *-ity* or *-ness* clearly cluster together. That is, bases that select *-ity* have different semantic properties than bases that select *-ness*. These results leave open whether Arndt-Lappe's phonological effect of the word's endings might emerge from the shared semantics of the respective bases, but as we saw in the general discussion, at least for the three clearly *-ity* or *-ness* leaning base-endings we also observe clear distinct patterns within the set of all bases considered in the study, indicating that here distributional semantics and morpheme/form based information go hand in hand. Besides establishing base semantics as a major factor in affix selection, study 1 also has shown that the distribution of the derivatives in the vector space shows a very similar clustering, in line with the idea that *-ity* and *-ness* induce similar meaning shifts. But as similar as these are, given the huge variation in resulting cosine similarities between bases and affixes and the fact that the affix itself remains as a significant predictor in the model this question needs further investigation. Study 2 used doublets to investigate the question of affix synonymy further. Since the doublets show only unsystematic differences, and since the frequencies of the derivatives play a significant role in accounting for this variation, one plausible explanation for any observed difference between pairs lies not in a systematic difference in meaning shifts brought about by the two affixes themselves but in effects of lexicalization.

In short, the semantic properties of the bases are a major factor in affix selection, while the two affixes themselves are synonyms to the extent that they have relatively similar effects on their bases.

References

Anshen, Frank and Mark Aronoff (1981). "Morphological Productivity and Phonological Transparency". In: *Canadian Journal of Linguistics/Revue canadienne de linguistique* 26.1, pp. 63–72. DOI: 10.1017/S0008413100023525.

Arndt-Lappe, Sabine (Nov. 2014). "Analogy in suffix rivalry: the case of English -ity and -ness". In: *English Language and Linguistics* 18 (03), pp. 497–548. ISSN: 1469-4379. DOI: 10.1017/S136067431400015X. URL: http://journals.cambridge.org/article_S136067431400015X.

Aronoff, Mark (1976). *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

Arora, Sanjeev, Wei Hu, and Pravesh K Kothari (2018). "An analysis of the t-SNE algorithm for data visualization". In: *Conference On Learning Theory*. PMLR, pp. 1455–1462.

Baayen, R. Harald and Antoinette Renouf (1996). "Chronicling the Times: Productive Lexical Innovations in an English Newspaper". In: *Language* 72.1, pp. 69–96. ISSN: 00978507, 15350665. URL: http://www.jstor.org/stable/416794 (visited on 03/15/2023).

Baeskow, Heike (2012). "-Ness and -ity: Phonological Exponents of n or Meaningful Nominalizers of Different Adjectival Domains?" In: *Journal of English Linguistics* 40.1, pp. 6–40. DOI: 10.1177/0075424211405156. URL: https://doi.org/10.1177/0075424211405156.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta (2009). "The WaCky wide web: a collection of very large linguistically processed web-crawled corpora". In: *Language Resources and Evaluation* 43.3, pp. 209–226. ISSN: 1574-0218. DOI: 10.1007/s10579-009-9081-4.

Baroni, Marco, Georgiana Dinu, and Germán Kruszewski (2014). "Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors". In: *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference* 1, pp. 238–247.

Bauer, Laurie, Rochelle Lieber, and Ingo Plag (2013). *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.

Biber, Douglas (1986). "Spoken and Written Textual Dimensions in English: Resolving the Contradictory Findings". In: *Language* 62.2, pp. 384–414. ISSN: 00978507, 15350665. URL: http://www.jstor.org/stable/414678 (visited on 02/12/2024).

— (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

— (1995). *Dimensions of register variation*. Cambridge: Cambridge University Press.

BNC (2007). *The British National Corpus, version 3 (BNC XML Edition)*. http://www.natcorp.ox.ac.uk/.

Boleda, Gemma (2020). "Distributional Semantics and Linguistic Theory". In: *Annual Review of Linguistics* 6, pp. 213–234.

Bonami, Olivier and Matías Guzmán Naranjo (2023). "Distributional evidence for derivational paradigms". In: *The Semantics of Derivational Morphology. Theory, Methods, Evidence*. Ed. by Sven Kotowski and Ingo Plag. Berlin, Boston: De Gruyter, pp. 219–258. ISBN: 9783111074917. DOI: doi:10.1515/9783111074917-008. URL: https://doi.org/10.1515/9783111074917-008.

Breheny, Patrick and Woodrow Burchett (2017). "Visualization of Regression Models Using visreg". In: *The R Journal* 9.2, pp. 56–71.

Cowie, Claire Susan. (1999). "Diachronic word-formation: a corpus-based study of derived nominalizations in the history of English." PhD thesis. Apollo - University of Cambridge Repository. DOI: 10.17863/CAM.16202. URL: https://www.repository.cam.ac.uk/handle/1810/251674.

Davies, Mark (2018). *The iWeb Corpus*. Available online at https://www.english-corpora.org/iWeb/.

Dixon, Robert M. W. (1982). *Where Have All the Adjectives Gone? and other Essays in Semantics and Syntax*. Berlin: Mouton de Gruyter.

Fellbaum, Christiane, ed. (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Kotowski, Sven and Ingo Plag, eds. (2023). *The Semantics of Derivational Morphology. Theory, Methods, Evidence*. Berlin, Boston: De Gruyter. ISBN: 9783111074917. DOI: doi:10.1515/9783111074917. URL: https://doi.org/10.1515/9783111074917.

Kotowski, Sven and Martin Schäfer (2023). "Quantifying semantic relatedness across base verbs and derivatives: English out-prefixation". In: *The Semantics of Derivational Morphology. Theory, Methods, Evidence*. Ed. by Sven Kotowski and Ingo Plag. Berlin, Boston: De Gruyter, pp. 177–218. ISBN: 9783111074917. DOI: doi:10.1515/9783111074917-007. URL: https://doi.org/10.1515/9783111074917-007.

Lindsay, Mark (2012). "Rival suffixes: synonymy, competition, and the emergence of productivity". In: *Morphology and the architecture of grammar: Proceedings of the 8th International Morphology Meeting*. Ed. by Angela Ralli, Geert Booij, Sergio Scalise, and Athanasios Karasimos. Vol. 8. University of Patras. Patras, pp. 192–203.

Maaten, Laurens van der and Geoffrey Hinton (2008). "Visualizing Data using t-SNE". In: *Journal of Machine Learning Research* 9, pp. 2579–2605.

Marchand, Hans (1969). *The Categories and Types of Present-Day English Word-Formation. A Synchronic-Diachronic Approach*. 2nd, completely revised and enlarged. München: C.H. Beck'sche Verlagsbuchhandlung.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (Jan. 2013). "Efficient Estimation of Word Representations in Vector Space". In: *ArXiv e-prints*.

Mikolov, Tomas, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, and Armand Joulin (2017). "Advances in Pre-Training Distributed Word Representations". In: *CoRR* abs/1712.09405. URL: http://arxiv.org/abs/1712.09405.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Plag, Ingo (2006). "The variability of compound stress in English: structural, semantic, and analogical factors". In: *English Language and Linguistics* 10.1, pp. 143–172.

Plag, Ingo, Christiane Dalton-Puffer, and R. Harald Baayen (1999). "Morphological productivity across speech and writing". In: *English Language & Linguistics* 3.2, pp. 209–228.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Riddle, Elizabeth M. (1985). "A historical perspective on the productivity of the suffixes -ness and -ity". In: *Historical Semantics— Historical Word-Formation*. Ed. by Jacek Fisiak. De Gruyter Mouton, pp. 435–462. DOI: doi:10.1515/9783110850178.435. URL: https://doi.org/10.1515/9783110850178.435.

Rodríguez-Puente, Paula, Tanja Säily, and Jukka Suomela (2022). "New methods for analysing diachronic suffix competition across registers: How -ity gained ground on -ness in Early Modern English". In: *International Journal of Corpus Linguistics* 27.4, pp. 506–528. DOI: https://doi.org/10.1075/ijcl.22014.rod. URL: https://www.jbe-platform.com/content/journals/10.1075/ijcl.22014.rod.

Sahlgren, Magnus (2006). "The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces". PhD thesis. Stockholm: Stockholm University.

Säily, Tanja (2011). "Variation in morphological productivity in the BNC: Sociolinguistic and methodological considerations". In: *Corpus Linguistics and Linguistic Theory*

7.1, pp. 119–141. DOI: doi:10.1515/cllt.2011.006. URL: https://doi.org/10.1515/cllt.2011.006.

Säily, Tanja (2018). “Chapter 12. Change or variation? Productivity of the suffixes "ness" and "ity"”. In: *Patterns of Change in 18th-century English*. Ed. by Terttu Nevalainen, Minna Palander-Collin, and Tanja Säily. John Benjamins, pp. 197–218. URL: https://www.jbe-platform.com/content/books/9789027263834-ahs.8.12sai.

Säily, Tanja and Jukka Suomela (2009). “Comparing type counts: The case of women, men and -ity in early English letters”. In: *Corpus Linguistics. Refinements and Reassessments*. Ed. by Antoinette Renouf and Andrew Kehoe. Leiden: Brill, pp. 87–109. DOI: https://doi.org/10.1163/9789042025981_007.

Schäfer, Martin (2023). “Splitting ly’s: Using word embeddings to distinguish derivation and inflection”. In: *The Semantics of Derivational Morphology. Theory, Methods, Evidence*. Ed. by Sven Kotowski and Ingo Plag. Berlin, Boston: De Gruyter, pp. 259–300. ISBN: 9783111074917. DOI: doi:10.1515/9783111074917-009. URL: https://doi.org/10.1515/9783111074917-009.

Shafaei-Bajestan, Elnaz, Masoumeh Moradipour-Tari, Peter Uhrig, and R. Harald Baayen (2022). *Semantic properties of English nominal pluralization: Insights from word embeddings*. DOI: 10.48550/ARXIV.2203.15424. URL: https://arxiv.org/abs/2203.15424.

Wood, S.N (2017). *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC.

Appendix: t-SNE across frequency bands

Below the t-SNE visualizations for the data from study 1 within the individual frequency bands, starting at the low end, the ultraLow band, and finishing with the highest band, the ultraHigh band. For ease of reference, the LDA scores already reported in table 5 are also given in the figure captions of the visualizations.

UltraLow frequency band

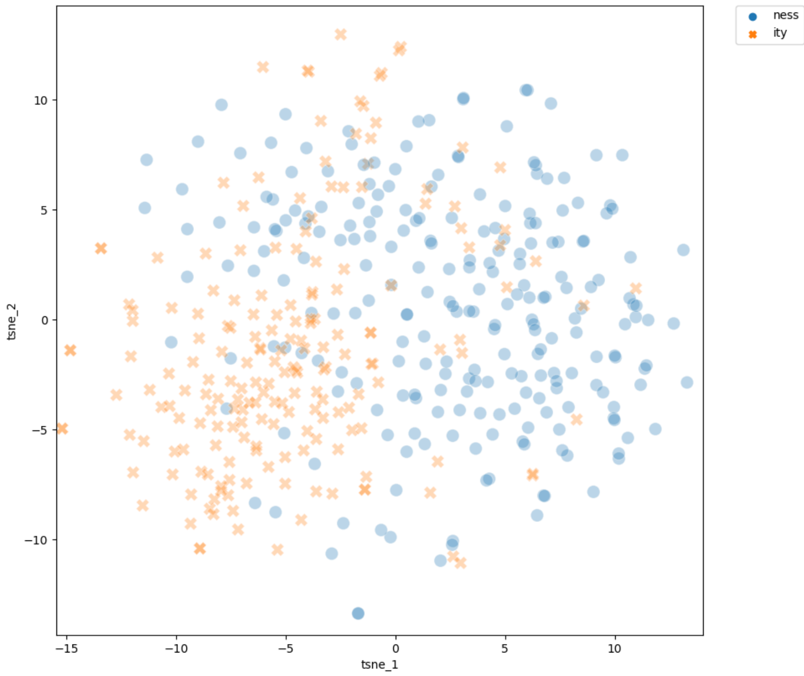

Figure 14: ADJ bases, ultra low frequency, no doublets. Corresponding mean weighted LDA scores: 0.686, standard deviation 0.061.

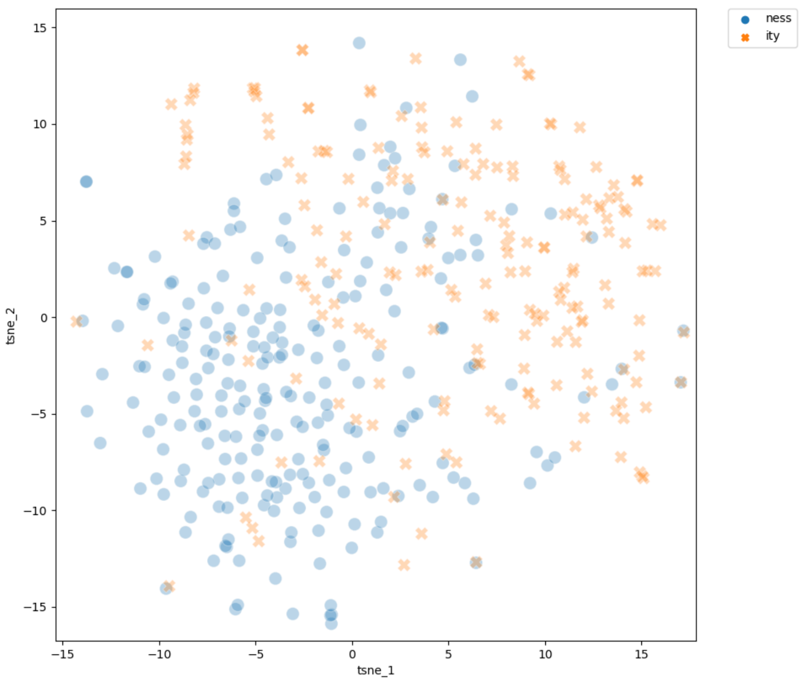

Figure 15: ADJ derivatives, ultra low frequency, no doublets. Corresponding mean weighted LDA scores: 0.693, standard deviation 0.066

Low frequency band

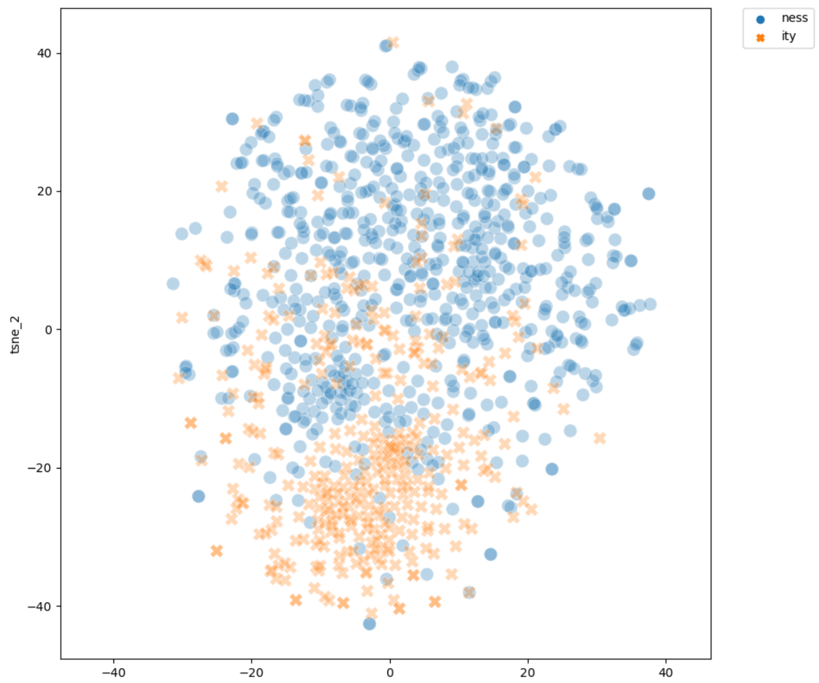

Figure 16: ADJ bases, low frequency, no doublets. Corresponding mean weighted LDA scores: 0.816, 0.04 standard deviation

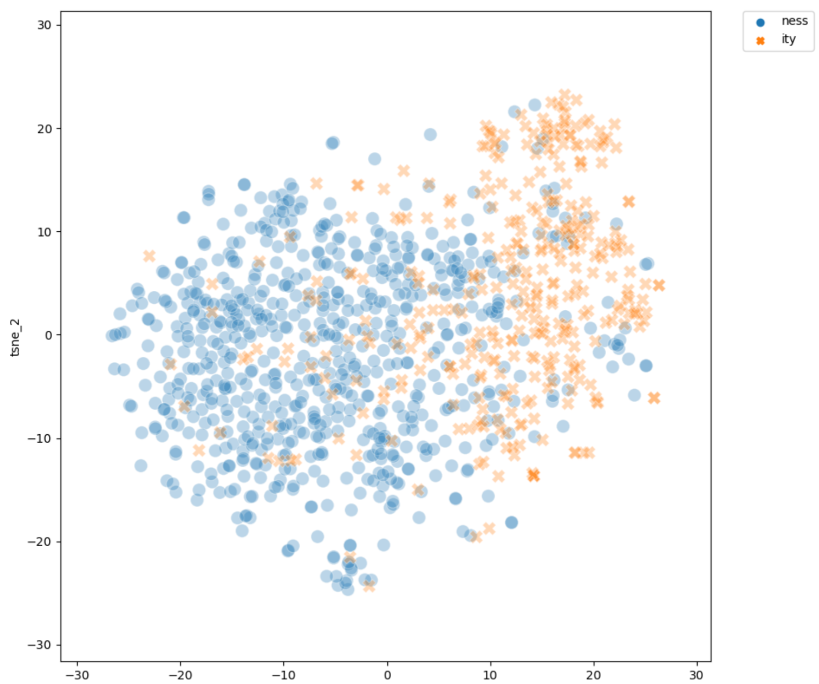

Figure 17: ADJ Derivatives, low frequency, no doublets. Corresponding mean weighted LDA scores: 0.821, 0.032 standard deviation

Mid frequency band

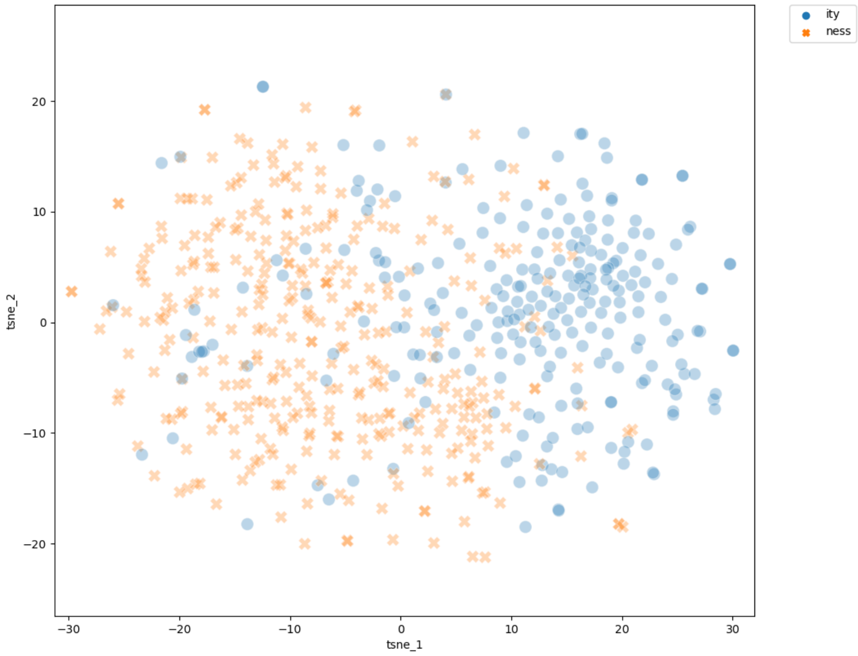

Figure 18: ADJ bases, mid frequency, no doublets. Corresponding mean weighted LDA scores: 0.753, 0.057 standard deviation

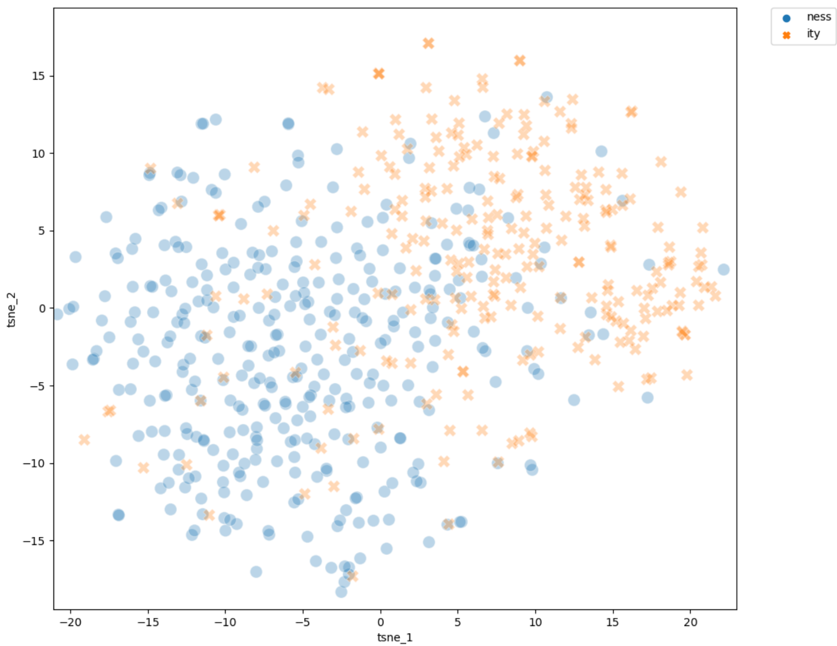

Figure 19: ADJ derivatives, mid frequency, no doublets. Corresponding mean weighted LDA scores: 0.747,0.057 standard deviation

MidHigh frequency band

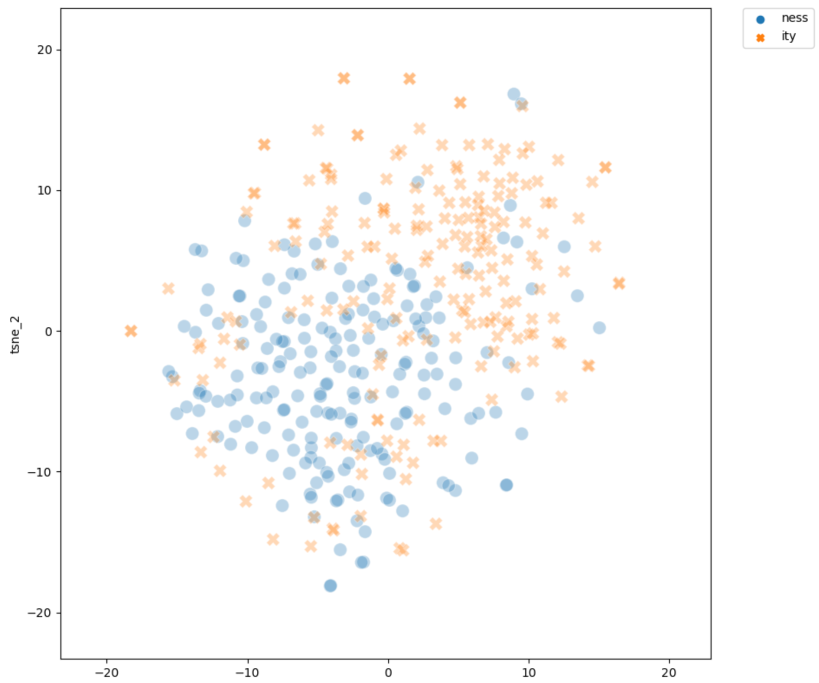

Figure 20: ADJ bases mid high frequency, no doublets. Corresponding mean weighted LDA scores: 0.665, 0.079 standard deviation

Figure 21: ADJ derivatives, mid high frequency, no doublets. Corresponding mean weighted LDA scores: 0.683, 0.08 standard deviation

High frequency band

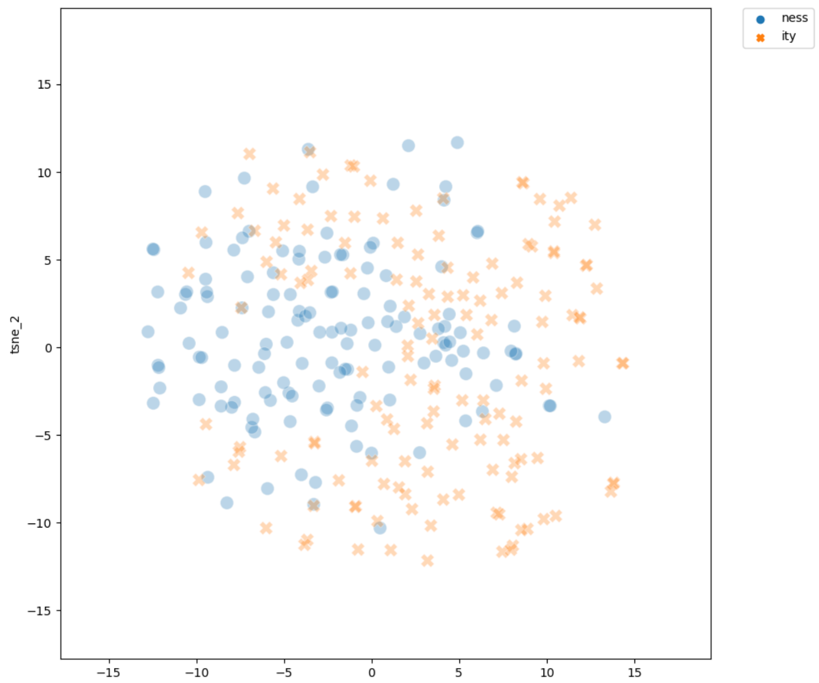

Figure 22: ADJ bases, high frequency, no doublets. Corresponding mean weighted LDA scores: 0.678, standard deviation 0.072

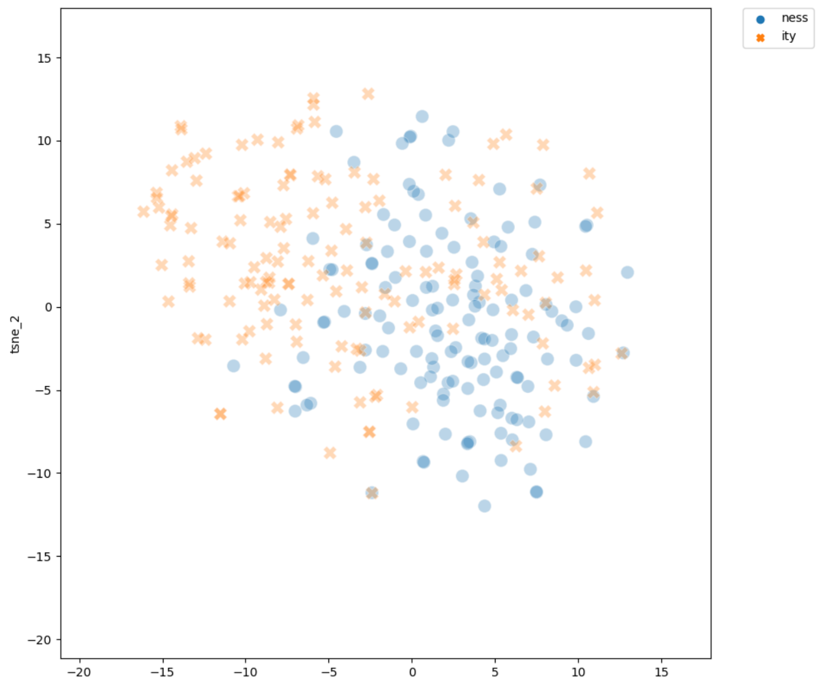

Figure 23: ADJ derivatives, high frequency, no doublets. Corresponding mean weighted LDA scores: 0.645, 0.094 standard deviation

SuperHigh frequency band

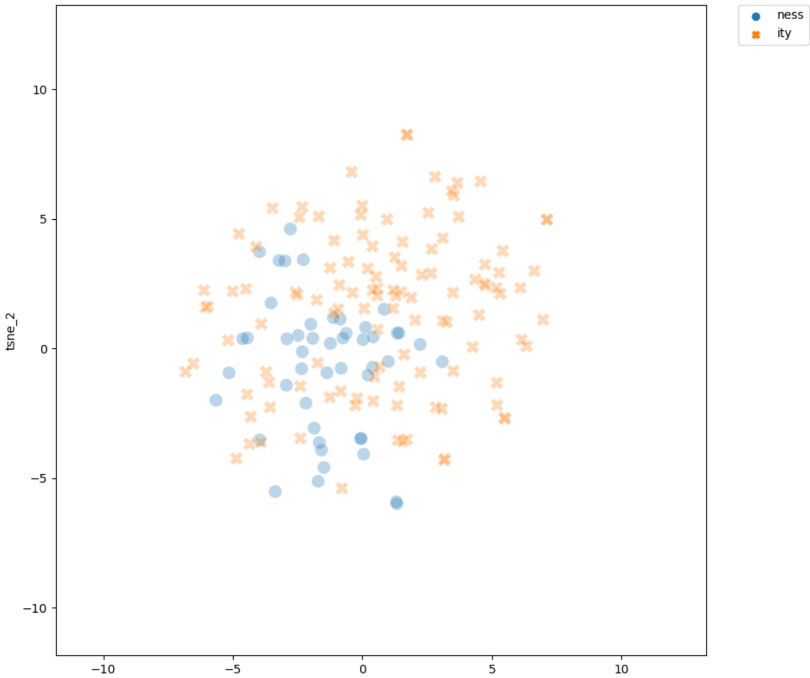

Figure 24: ADJ bases, super high frequency, no doublets. Corresponding mean weighted LDA scores: 0.756, standard deviation 0.075

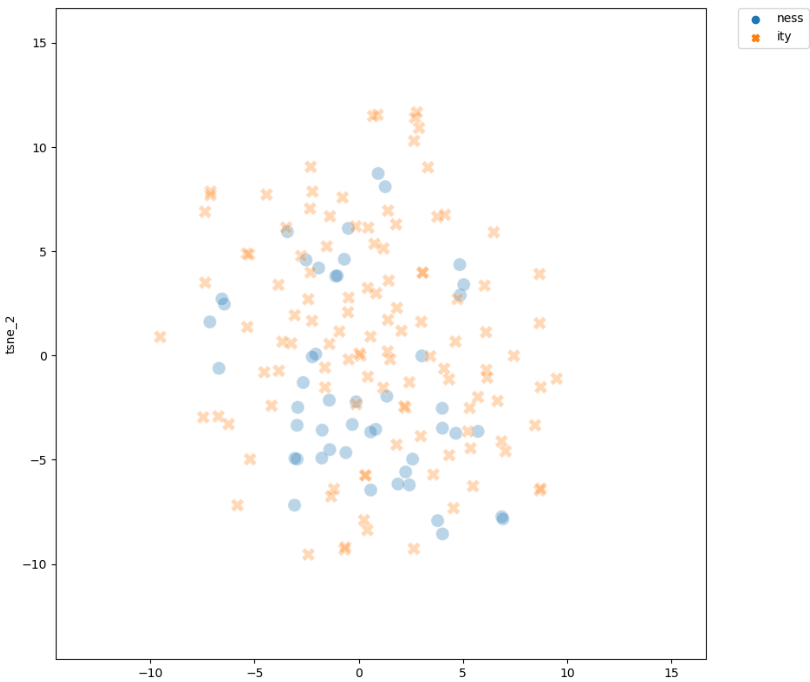

Figure 25: ADJ derivatives, super high frequency, no doublets. Corresponding mean weighted LDA scores: 0.777, standard deviation 0.102

UltraHigh frequency band

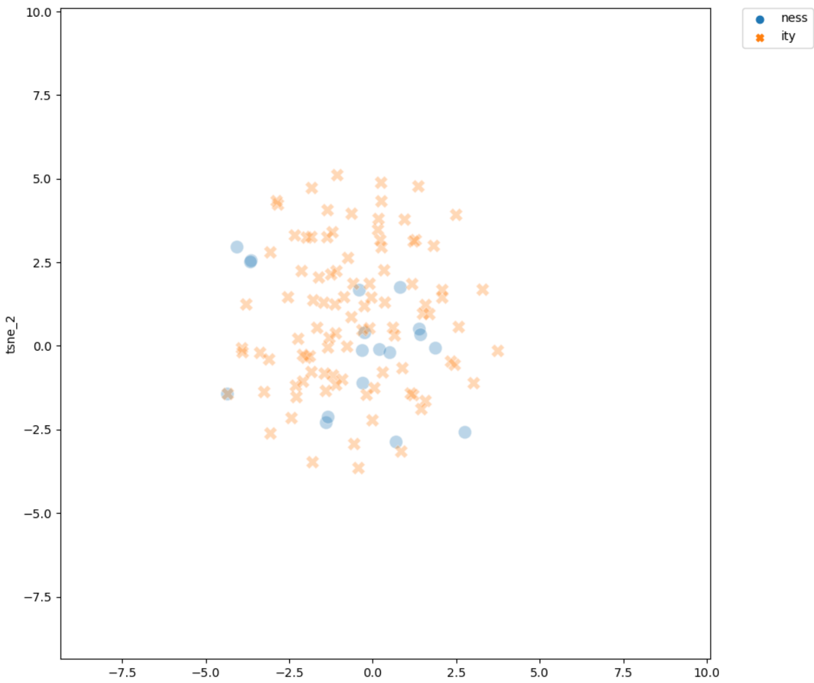

Figure 26: ADJ Bases ultra high No Doublets. Corresponding mean weighted LDA scores: 0.785, 0.09 standard deviation

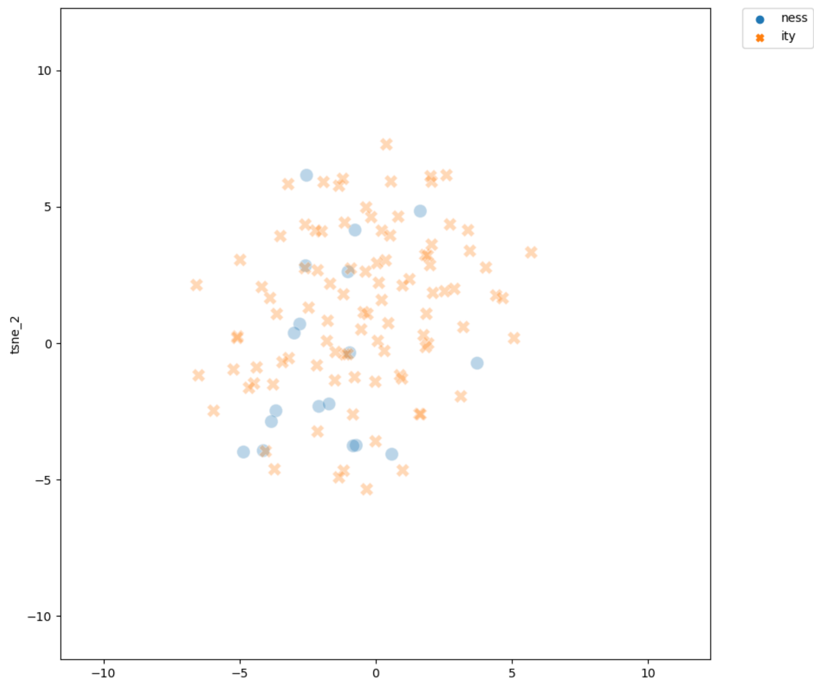

Figure 27: ADJ derivatives, ultra high frequency, no doublets. Corresponding mean weighted LDA scores: 0.805, 0.087 standard deviation