

# Disambiguating Grammatical Number and Gender With BERT

Annegret Janzso

Saarland University

Department of Language Science and Technology

66123 Saarbrücken, Germany

s8anjanz@stud.uni-saarland.de

## Abstract

Accurately dealing with any type of ambiguity is a major task in Natural Language Processing, with great advances recently reached due to the development of context dependent language models and the use of word or sentence embeddings. In this context, our work aimed at determining how the popular language representation model BERT handles ambiguity of nouns in grammatical number and gender in different languages. This work shows that models trained on one specific language achieve better results for the disambiguation process than multilingual models. Also, ambiguity is generally better dealt with in grammatical number than it is in grammatical gender, reaching greater distance values from one to another in direct comparisons of individual word sense embeddings. The overall results show also that the amount of data needed for training monolingual models as well as application should not be underestimated.

## 1 Introduction

A challenge in Natural Language Processing (NLP) resides in the accurate automatic sense disambiguation of words and phrases. An often cited example of ambiguity is the one between *bank* ("An institution where one can place and borrow money and take care of financial affairs.") and *bank* ("An edge of river, lake, or other watercourse.") in English.<sup>1</sup> While a person is able to get the right meaning of the word from context, the same skill is now expected from contextual language models.

The goal of our work, however, is to find out if and how well contextual meaning representations can handle sense ambiguities in grammatical gender and number. Further, this paper aims to show differences in disambiguation in between different

<sup>1</sup>Definitions taken from Wiktionary: <https://en.wiktionary.org/wiki/bank>

languages, ambiguity types, and pre-existing models. For these tasks, German and Spanish are used for gender and number ambiguities, and English for number ambiguities only (see section 3). Ambiguity in grammatical number is similar to the example mentioned above, as the plural form of a word might also mean something different from the mere quantity. In this case, the Language Model (LM) needs to disambiguate the meaning of plurals considering only the current context. For grammatical gender ambiguity - that is, words that can occur in more than one gender, with their meaning dependent on that factor - additional cues to the current meaning can be found in words that are connected to the currently observed ambiguous word, such as accordingly gendered determiners or adjectives.

This specific skill of disambiguation of LMs is tested on BERT word embeddings (see section 4.2), based on data of three different languages: German (section 3.1) and Spanish (section 3.2), which contain both types of ambiguity, and English (section 3.3), which contains only ambiguity in grammatical number, but has been included due to the large amount of available data for testing (see section 4.1).

## 2 Related Work

While existing work on ambiguity mostly focuses on a more general definition of the term, grammatical gender and number have mostly been left aside. However, there has been research on gender bias in Language Models. Bartl et al. (2020) found a gender-bias for English in BERT similar to the gender-bias occurring with names of professions. For German, this effect has been found to be less strong, due to morphological gender markings. Seeing how a language without grammatical gender shows a higher gender bias than a language with such, this topic appears to be rather drifting

in a direction of biological gender and social prejudice connected with gender, which, according to [Bartl et al. \(2020\)](#), can be solved by grammatical gender in form of morphological markings, as this occurs in German. These results show a possibility to search for a gender bias in the disambiguation process of BERT for languages such as German or Spanish.

Concerning ambiguity in grammatical number, [Gromann and Declerck \(2019\)](#) analyses how plural entries included in Princeton WordNet ([Princeton University, 2010](#)) are encoded in word embeddings using Word2Vec ([Mikolov et al., 2013](#)) and GloVe ([Pennington et al., 2014](#)), thereby showing the convergence between distributed representations of plurals bearing a specific sense, as well as their symbolic representation proposed in WordNet. They bring up examples like *people - peoples*: Both words occur in plural form. However, *people* can be treated as a singular term, meaning 'the body of citizens of a state or country' (Definition given by Princeton Wordnet ([Princeton University, 2010](#))). *Peoples* can be used as the plural of *people*, but can also refer to 'the human beings of a particular nation or community or ethnic group' (Definition given by Princeton Wordnet ([Princeton University, 2010](#))). This shows that:

1. the singular and plural form of a word don't necessarily share all their meanings and
2. a word can function as a plural form for some senses, but as a singular form for other senses

This shows further that there is a lot of possibility for ambiguity in grammatical number, making it just as relevant as grammatical gender in a task of disambiguation.

Considering sense already during training, instead of researching on existing models, has been shown to be a promising approach (see [Blevins and Zettlemoyer \(2020\)](#), [Levine et al. \(2019\)](#)) for English data. These approaches add information on senses or glosses to the training process, thereby considering this additional knowledge in all applicational usage.

Morphological approaches have also shown promising results in English, German ([Cotterell and Schütze, 2018](#)) and Hebrew ([Avraham and Goldberg, 2017](#)). Future research in that direction will show if these approaches can improve disambiguation of those ambiguities that contain relevant morphological cues (see [Corbett \(2007\)](#)).

Observations of morphological content of sentences, including grammatical number and gender, have shown that BERT achieves better results for morphologically simple languages, in comparison to morphologically more complex languages such as German or Russian (see [Edmiston \(2020\)](#)). This work shows that BERT is able to detect and disambiguate such morphological content based on contextual cues, but does not achieve human-like results.

### 3 Ambiguity in Grammatical Gender and Number

The present work takes focus on ambiguity occurring in grammatical number and gender.

Ambiguity in grammatical number occurs if a word **w** has a plural **p** that has a standing meaning of its own, that is not only expressing a quantity associated with the word. An example in English is given by the word *glass*; its plural *glasses* can either mean *multiple items of the object named glass*, or *an object that is worn on one's face to increase eyesight*. While both meanings of *glasses* appear to be etymologically related, they do mean different things, and one meaning is independent of the meaning of the singular word *glass*.

Ambiguity in grammatical gender is given when a word **w** is ambiguous, while **w** can be assigned to at two or more different grammatical genders **g** and **g'**, and the standing meaning is dependent on the assigned gender. A German example is displayed in Table 1.

Gender	Word	Meaning
<b>masculine</b>	[der] Kiefer	[the] jaw
<b>feminine</b>	[die] Kiefer	[the] pine tree

Table 1: Differences in gender for the German word *Kiefer* with a change in meaning.

This work deals with grammatical number ambiguity in English, German, and Spanish, and grammatical gender ambiguity in German and Spanish. The following sections handle languages-specific details related to such ambiguities.

#### 3.1 German

The German language contains both grammatical gender and grammatical number ambiguity.

Grammatical number ambiguity often occurs when two words **w** and **p** that look like they are

forming a singular-plural pair have different meanings. In fact the two words are not forming such a pair, but are the cases of a *singulare tantum* and a case of *plurale tantum*, as can be seen in Table 2

Number	Word	Meaning
<b>Singular</b>	Schuld	guilt
<b>Plural</b>	Schulden	debt

Table 2: Different meanings depending on the grammatical number in German.

Here, **w** equals *Schuld* and **p** equals *Schulden*. The morphological *-en* ending is often added to German nouns to create a plural, making **w** and **p** appear as a singular-plural pair, but they are a pair of *singulare tantum* and *plurale tantum*, each with their particular meaning. German has three different grammatical genders: masculine, feminine, and neutral. When the grammatical gender of a word changes, so do its article and adjectives in the sentence, which are a good clue to gender detection, and therefore possibly also for meaning disambiguation. Grammatical gender ambiguity, as described in section 3, occurs when an ambiguous word **w** can be assigned two or more grammatical genders, and is assigned a different meaning for each. The last part is important, because some words can be used in more than one grammatical gender, without having a different meaning, as shown in Table 3. An example of actual gender ambiguity in German is shown in Table 1.

Gender	Word	Meaning
<b>masculine</b>	[der] Paprika	[the] bell pepper
<b>feminine</b>	[die] Paprika	[the] bell pepper

Table 3: Differences in gender for the German word *Paprika* without a change in meaning.

### 3.2 Spanish

Just like in German, it is possible to observe both types of ambiguity defined above in section 3 in Spanish. Grammatical number ambiguity usually occurs when the plural **p** of a word **w** is ambiguous, and one meaning has no singular term, or one that is different to **w**. An example can be found in the word *esposa*, as shown in Table 4.

Number	Word	Meaning 1	Meaning 2
<b>Singular</b>	esposa	wife	
<b>Plural</b>	esposas	wives	handcuffs

Table 4: Change in meaning depending on grammatical number in Spanish.

For grammatical gender ambiguity, there are two grammatical genders in Spanish: masculine and feminine. Other than that, it is defined the same way it is for German in section 3.1. An example can be the word *cólera*, as shown in Table 5

Gender	Word	Meaning
<b>masculine</b>	[el] cólera	[the] cholera
<b>feminine</b>	[la] cólera	[the] anger

Table 5: Change in meaning depending on grammatical gender in Spanish.

### 3.3 English

Since English is (in its majority) genderless, it is enough to state that the present work in English deals only with ambiguity in grammatical number.

Number ambiguity in English usually occurs when the plural **p** of a word **w** is ambiguous, and one meaning has no singular term, or one that is different to **w**. Other cases - such as a plural **p** being able to used as a singular term **w**, which can be assigned another plural term **p'** - have been observed in other works (e.g. Gromann and Declerck (2019)) with examples such as *person* for **w**, *people* for **p**, and *peoples* for **p'**. These types of occurrences, however, have not been included in this work.

An example on grammatical number ambiguity in English is described in section 3 with the terms *glass* and *glasses*.

## 4 Methodology

In the following sections, the methods used to compute and compare embedding vectors are described. Section 4.1 describes the data and its necessary information. Section 4.2 lists the pretrained BERT models that have been used to retrieve results, which then have been evaluated using the methods described in section 4.3. The actual results and evaluation outcomes are later described in section 5.

### 4.1 Data

**Wiktionary** (Wikimedia, 2021b) is a free online dictionary, created by the Wikimedia Foundation

(Wikimedia, 2021a), providing a range of XML data dumps. It contains information of over 170 languages, and provides detailed information on every word, including Pronunciation, Etymology, Translations, and of course grammatical number and gender, as well as according meanings and example sentences, when any information of a type is available.

Our work uses three so-called Wiktionary dumps (one per language) from July 2021. Each dump has been parsed by taking individual changes in between dumps into account, and information relevant to the task (title, possible senses, example sentences, possible plurals, grammatical number, grammatical gender) have been filtered out for all noun entries. Some entries of the dump of one language are written for words in another language, e.g. there might be an entry for a Spanish word in the English dump. When this happened for one of the three languages used in this work, these entries have also been parsed and saved in additional files, so they could later be used as additional data for their languages.

Entries without the necessary information (e.g. missing example sentences for context) have been discarded, if there was no other way to retrieve the needed information from another dump.

There are significant differences in the size of each dataset, which have been further adjusted to eliminate entries irrelevant for the tasks at hand - such as word types other than nouns, proper names, and entries without any of the observed ambiguities.

To gather additional example sentences for English entry data, NLTK WordNet (Princeton University, 2010) has been used. For any English noun without an example sentence, the word was looked up in WordNet and all available senses and example sentences have been added to the collected data.<sup>2</sup>

An idea of how much data is available to use per language can be given by an overview of iterations used when computing word embeddings (see Table 12 in the appendix). For each language, only a few examples will be used for representation purposes in this paper. Unfortunately, only a small amount of data is available for Spanish, which means only a small amount of possible comparisons. Overall

<sup>2</sup>Some example sentences in WordNet contain a synonym of a word instead of the word that is needed. In these cases, the synonym has been replaced by the word in question. This may lead to grammatical errors within the sentence, which has been ignored.

there was not enough data on Spanish to compute vector embedding distances, which is why this part is not included in the results in section 5.

## 4.2 BERT

BERT (Devlin et al., 2018) is a bidirectional language model and is by now a commonly used tool for NLP. The goal is to test its ability of disambiguation with regard to grammatical gender and number, as a representative of current methods, by computing word embeddings of ambiguous words in specific contexts and specific meanings. Also, the many pre-existing models available for BERT eliminate the need to train any models for the purposes of our work.

Four pre-trained models have been used here:

- BERT base uncased (Devlin et al., 2018)
- BERT base multilingual uncased (Devlin et al., 2018)
- BERT base German uncased<sup>3</sup>
- BERT base Spanish wwm uncased (Cañete et al., 2020)

By now, for every language, there is one language-specific pre-trained model; the multilingual model includes all three languages, among others.

## 4.3 Evaluation Methods

To compare the word embedding vectors created by BERT, three methods for computing distances in vector space have been used: **Cosine**, **Euclidean**, and **Manhattan** distance.

Cosine Distance represents the angle between the origins of two vectors. If cosine distance equals 0, the vectors are identical. The higher the distance value, the more different the vectors are from each other. Under the assumption that words with the same meaning will also have the same or similar embedding vectors, and words with distinguished meanings shouldn't, this means that if BERT can properly disambiguate words in this task, the cosine distance should be rather large for ambiguous words with different meanings. If, however, cosine distance is small and close to 0, this would translate as BERT not properly disambiguating the word in its multiple meanings.

<sup>3</sup><https://huggingface.co/dbmdz/bert-base-german-uncased>

Given that computing distance by angle in a multi-dimensional vector space might be imprecise in some cases, additional methods have been used. Euclidean Distance portrays the shortest distance from one point in vector space to another, regardless of dimensionality. Similar to Euclidean Distance, a value of 0 means the embedding vectors appear to be identical, whereas larger values portray the opposite.

A third method used is Manhattan Distance, which, given a rectangle formed by the two points in vector space, is computed by the sum of the lengths of this rectangle. The resulting values are to be treated similarly to the values of Cosine and Euclidean Distance.

## 5 Results

Distances are compared between ambiguity-types: Comparing results of number-ambiguity to gender-ambiguity shows which of the two is processed better. Comparison of either to 'classic' ambiguity, without a change in number or gender, shows if there is any improvement due to additional contextual cues. Distances between two different, non-ambiguous words are treated as baseline: When ambiguous words have a similar distance as non-ambiguous words, the disambiguation process was successful. These results are presented in section 5.1.

Differences occurring due to the use of different models (multilingual or language specific) show the relevance of being specific to one language. These comparisons can be found in section 5.2.

Due to a huge lack of data, especially in example sentences that would be required to provide context, no distances could be computed on Spanish data. Computed data on Spanish ambiguity in grammatical gender only contained words where the sense stays the same with either possibility of gender. Computed data on Spanish ambiguity in grammatical number only contained words in singular, but none in plural. Therefore, this language is left out in further comparisons. For all other languages, the words and senses that have been used for comparisons in this paper can be found in Table 13 in the appendix.

Cosine Distances have been left out of the tables for results, as most were computed to be 0.0. Those with different values showed similar tendencies to the results of euclidean and manhattan distance, which can be compared better.

<b>Sg</b>	<b>Pl</b>	<b>Euc</b>	<b>Man</b>
kitchening	kitchenings	1.54e <sup>+75</sup>	7.68e <sup>+75</sup>
wood	woods	0.321458	8.897009
gen	gens	1.55e <sup>+75</sup>	7.74e <sup>+75</sup>
<b>W</b>	<b>W'</b>		
kitchening	wood	1.54e <sup>+75</sup>	7.69e <sup>+75</sup>
kitchening	gen	1.54e <sup>+75</sup>	7.69e <sup>+75</sup>
wood	gen	0.0	8.550618

Table 6: Results for ambiguity in grammatical number in English in comparison to words without relation of number (monolingual model)

Sg=Singular, Pl=Plural, Euc=Euclidean Distance, Man=Manhattan Distance, W=Word

<b>Sg</b>	<b>Pl</b>	<b>Euc</b>	<b>Man</b>
kitchening	kitchenings	0.213219	5.906484
wood	woods	0.072258	2.002483
gen	gens	1.51e <sup>+75</sup>	7.54e <sup>+75</sup>
<b>W</b>	<b>W'</b>		
kitchening	wood	0.098116	2.719078
kitchening	gen	0.106438	2.949704
wood	gen	0.008322	0.230625

Table 7: Results for ambiguity in grammatical number in English in comparison to words without relation of number (multilingual model)

Sg=Singular, Pl=Plural, Euc=Euclidean Distance, Man=Manhattan Distance, W=Word

<b>W</b>	<b>G</b>	<b>G'</b>	<b>Euc</b>	<b>Man</b>
Band	m	n	0.209174	5.799105
Gehalt	n	m	0.086309	2.403371
Leiter	f	m	0.194705	5.395817
<b>W</b>	<b>W'</b>	<b>G</b>	<b>Euc</b>	<b>Man</b>
Band	Leiter	m	0.219066	3.717639
Band	Gehalt	n	0.160602	4.445368
Leiter	Gehalt	m	0.144773	4.006747

Table 8: Results for ambiguity in grammatical gender in German in comparison to words without relation of gender (monolingual model)

W=Word, G=Gender, Euc=Euclidean Distance, Man=Manhattan Distance

W	G	G'	Euc	Man
Band	m	n	0.248292	6.871943
Gehalt	n	m	0.198916	5.508927
Leiter	f	m	0.008563	0.237300
W	W'	G	Euc	Man
Band	Leiter	m	0.204392	5.665711
Band	Gehalt	n	0.092686	2.568588
Leiter	Gehalt	m	0.062330	1.742512

Table 9: Results for ambiguity in grammatical gender in German in comparison to words without relation of gender (multilingual model)

W=Word, G=Gender, Euc=Euclidean Distance, Man=Manhattan Distance

Sg	Pl	Euc	Man
Schuld	Schulden	0.124265	3.444058
Sasse	Sassen	0.093512	2.590606
Barre	Barren	0.0	1.315709
W	W'	Euc	Man
Schuld	Sasse	0.173317	4.812258
Schuld	Barre	$1.48e^{+75}$	$7.27e^{+75}$
Sasse	Barre	$1.48e^{+75}$	$7.27e^{+75}$

Table 10: Results for ambiguity in grammatical number in German in comparison to words without relation of number (monolingual model)

Sg=Singular, Pl=Plural, Euc=Euclidean Distance, Man=Manhattan Distance

Sg	Pl	Euc	Man
Schuld	Schulden	0.066630	1.846514
Sasse	Sassen	0.167143	4.638568
Barre	Barren	0.017129	0.474700
W	W'	Euc	Man
Schuld	Sasse	0.103715	2.899168
Schuld	Barre	0.034782	0.963905
Sasse	Barre	0.138497	3.853023

Table 11: Results for ambiguity in grammatical number in German in comparison to words without relation of number (multilingual model)

Sg=Singular, Pl=Plural, Euc=Euclidean Distance, Man=Manhattan Distance

## 5.1 Number and Gender Ambiguity

Cosine distance, euclidean distance, and manhattan distance have been computed for word pairs on grammatical gender for German and grammatical number for English and German. Comparisons of individual words out of those pairs, which do not contain such a relation of grammatical gender or number (respectively), are used as a baseline.

For ambiguity in grammatical number in English, results are found in Tables 6 and 7 for mono- and multilingual models, respectively. Results regarding the differences between the two models can be found in section 5.2. The results show similar results for ambiguous and non-ambiguous word pairs.

For ambiguity in grammatical gender in German, results are found in Tables 8 and 9 for mono- and multilingual models, respectively. The results show similar results for ambiguous and non-ambiguous word pairs, however the ambiguous word pairs achieve slightly larger distances in some cases.

For ambiguity in grammatical number in German, results are found in Tables 10 and 11 for mono- and multilingual models, respectively. Just like for ambiguity in grammatical gender, results for ambiguous and non-ambiguous word pairs are similar, but slightly larger distances have been computed for non-ambiguous word pairs, with exception of the word pair of *Schuld* and *Schulden*, which overall achieved rather high distance values.

## 5.2 Monolingual and Multilingual Models

For ambiguity in grammatical number in English, results can be found in Table 6 and Table 7 for the mono- and multilingual models, respectively. Overall, the monolingual model was able to disambiguate better than the multilingual model, resulting in greater distances for both methods of distance computations. Distances are greater in the monolingual model for the ambiguous as well as the non-ambiguous case.

For ambiguity in grammatical gender in German, results can be found in Table 8 and Table 9 for the mono- and multilingual models, respectively. Overall, the models achieved very similar results. Both models perform slightly better on ambiguous than on non-ambiguous data.

For ambiguity in grammatical number in German, results can be found in Table 10 and 11 for the mono- and multilingual models, respectively. Overall, the monolingual model was able to disam-

biguate better than the multilingual model, resulting in greater distances for both methods of distance computations. Distances are slightly greater in the monolingual model for the ambiguous case, but much greater for the non-ambiguous case.

## 6 Discussion

Given the results found in section 5.1, BERT can disambiguate ambiguity in grammatical number in English similarly well as non-ambiguous word pairs. The same goes for ambiguity in grammatical gender in German. Overall, ambiguity in grammatical number in German could also be disambiguated well, but not as well as for the same type of data in English. A reason for this outcome might be more available data to compute word embeddings from or possible differences in amounts of data per language during training of the individual models.

Ambiguity in grammatical number could be disambiguated better than for grammatical gender, which is possibly due to additional morphological cues within the word, or in some cases a more notable difference within the provided context.

Given the results found in section 5.2, the monolingual model bert-base-uncased outperformed the multilingual model bert-base-multilingual on English data. Concerning ambiguity in grammatical gender in German, both models performed similarly well on both ambiguous and non-ambiguous data. In the case of ambiguity in grammatical number in German, the monolingual model outperformed the multilingual model, achieving slightly greater results for ambiguous data, and much greater results for non-ambiguous data. Overall did the multilingual model perform better on German data, especially on ambiguity in grammatical gender, while it performs similarly well on ambiguity in grammatical number for both languages. The English monolingual model achieved much larger distances overall in comparison to the German monolingual model.

Using monolingual models is rewarded with better disambiguation, showing that it is well worth the time invested in the creation of such models, as the multilingual counterpart does well, but by far not as good. It is, however, a great way to include minority languages with not enough data to create individual language models. For languages like English and German, however, monolingual models appear to be the way to go, at least for disambiguation tasks.

## 7 Conclusion

This work shows that basic BERT models for English are better adapt to specific types of ambiguity than those for other languages like German, shown by the greater distances reached in the given evaluation methods. However, the multilingual model did perform better on German data than on English data, showing that a change of focus can improve results on languages with less data, but overall does not perform as well as language specific models.

This work shows a need for well-trained monolingual models, which appear to provide a better possibility of focus on disambiguation tasks. Further, available data such as provided by Wiktionary (Wikimedia, 2021b) needs to increase.

Interesting expansions of this work could come with approaches such as SenseBERT (Levine et al., 2019) - a BERT model that is trained on WordNet (Princeton University, 2010) supersenses - which might be able to achieve better results in disambiguation, and goes along with the findings of Blevins and Zettlemoyer (2020). Switching from word level to a morphological level could also bring up interesting results (Avraham and Goldberg (2017), Cotterell and Schütze (2018)).

## Acknowledgements

This work has been partly supported by the Horizon 2020 research and innovation programme with the project Prêt-à-LLOD (grant agreement no. 825182) at DFKI Saarbrücken. I thank Thierry Declerck, Josef van Genabith, and Lucia Donatelli for their helpful advice.

## References

- Oded Avraham and Yoav Goldberg. 2017. [The interplay of semantics and morphology in word embeddings](#). *CoRR*, abs/1704.01938.
- Marion Bartl, Malvina Nissim, and Albert Gatt. 2020. [Unmasking contextual stereotypes: Measuring and mitigating bert’s gender bias](#). *CoRR*, abs/2010.14534.
- Terra Blevins and Luke Zettlemoyer. 2020. [Moving down the long tail of word sense disambiguation with gloss-informed biencoders](#). *CoRR*, abs/2005.02590.
- José Cañete, Gabriel Chaperon, Rodrigo Fuentes, Joun-Hui Ho, Hojin Kang, and Jorge Pérez. 2020. [Spanish pre-trained bert model and evaluation data](#). *PMLADC at ICLR 2020*.

- Greville G. Corbett. 2007. Canonical typology, suppletion, and possible words. *Language*, 83(1):8–42.
- Ryan Cotterell and Hinrich Schütze. 2018. Joint semantic synthesis and morphological analysis of the derived word. *Trans. Assoc. Comput. Linguistics*, 6:33–48.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.
- Daniel Edmiston. 2020. A systematic analysis of morphological content in BERT models for multiple languages. *CoRR*, abs/2004.03032.
- Dagmar Gromann and Thierry Declerck. 2019. Towards the Detection and Formal Representation of Semantic Shifts in Inflectional Morphology. In *2nd Conference on Language, Data and Knowledge (LDK 2019)*, volume 70 of *OpenAccess Series in Informatics (OASIs)*, pages 21:1–21:15, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- Yoav Levine, Barak Lenz, Or Dagan, Dan Padnos, Or Sharir, Shai Shalev-Shwartz, Amnon Shashua, and Yoav Shoham. 2019. Sensebert: Driving some sense into BERT. *CoRR*, abs/1908.05646.
- Tomás Mikolov, Quoc V. Le, and Ilya Sutskever. 2013. Exploiting similarities among languages for machine translation. *CoRR*, abs/1309.4168.
- Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. GloVe : Global Vectors for Word Representation. *Proceedings of the 2014 Conference on EMNLP*, pages 1532–1543.
- Princeton University. 2010. About wordnet. <https://wordnet.princeton.edu/>.
- Wikimedia. 2021a. Wikimedia foundation. <https://wikimediafoundation.org/>.
- Wikimedia. 2021b. Wiktionary. <https://www.wiktionary.org/>.



## A Appendix

Language	Type	Count
<b>German</b>	Number	1606
	Gender	1672
	Neither	99250
<b>Spanish</b>	Number	22
	Gender	42
	Neither	6233
<b>English</b>	Number	151610
	Neither	203077

Table 12: Amount of iterations per language and ambiguity type used to compute word embeddings.

<b>English</b>		
Word	Number	Sense
kitchening	Sg	food preparation
kitchenings	Pl	scraps of food waste [...]
wood	Sg	a peckerwood
woods	Pl	A dense collection of trees [...]
gen	Sg	a specific version of something in chronological sequence
gens	Pl	a tribal subgroup [...]
<b>German</b>		
Word	Gender	Sense
Band	m	singular book that is part of a series [...]
	n	the [...] ribbon that is stretched over a finish line
Gehalt	n	amount of money that is [...] paid to workers
	m	the mental [...] content of a creative work
Leiter	f	stitch in knitted [...] wares [...] that got loose [...]
	m	material that lets energy [...] flow through
Word	Number	Sense
Schuld	Sg	moral misbehaviour [...]
Schulden	Pl	the entire obligations (passiva) of a person [...]
Sasse	Sg	hunting language: a flat trough, used by rabbits as a spot for rest and hiding
Sassen	Pl	historical landmark in east middle europe [...]
Barre	Sg	area of shallow water
Barren	Pl	metal poured in form

Table 13: Words and senses used for comparisons in this paper; the (translated) senses come from the according entries in Wiktionary