Distributional semantics and the conceptual foundations of verb meaning: How neural word embeddings memorize the unaccusative hypothesis

Tillmann Pross

Abstract I investigate whether and how neural word embeddings can be understood to encode not only idiosyncratic aspects of word meaning but also the kind of general and abstract concepts that are central to theoretical approaches of lexical semantics. To this end, I compute the difference between general-purpose embeddings of intransitive verbs, and task-specific embeddings of the same verbs that capture their similarity according to the unaccusative hypothesis. I show that the difference that retraining makes is neither trivial nor random but captures surpisingly well the cues for unergativity and unaccusativity that have been proposed in the theoretical literature. The study presented thus suggests that word embeddings may provide a novel and empirically grounded perspective on the conceptual underpinnings of verb meaning.

Keywords unaccusativity · word embedding · thematic role · lexical semantics

T. Pross, University of Stuttgart, tillmann.pross@ims.uni-stuttgart.de

In Christopher Pinon & Laurent Roussarie (eds.), Empirical Issues in Syntax and Semantics 13, 81–107. Paris: CSSP. http://www.cssp.cnrs.fr/eiss13/

1 Introduction

A theory of lexical representation is key to compositional theories of the meaning of phrases and sentences. One of the main challenges for the development of such a theory is the identification and specification of the primitive meaning components out of which verb meanings are constructed. While such a theory of the primitive components of verb meaning is of central importance to the development of a general theory of the lexical semantics of verbs – "[t]he important theoretical construct is the notion of meaning component, not the notion of verb class" (Levin 1993: 18) – the determination of the components of verb meaning is a central, yet unsolved research problem in both theoretical and computational approaches to verb meaning. By means of a case study on the meaning com-

ponents of intransitive verbs, this paper ties in with recent work (such as Asher et al. 2016; McNally & Boleda 2017; Pross et al. 2017, see e.g. Boleda 2020 for a review of the current state of the art) that aims to show that this situation is to the detriment of both theoretical and computational approaches to lexical semantics. Accordingly, the goal of the present paper is to show that striving for a combination of theoretical and computational approaches to lexical semantics is a natural way to deal with problems which can not be solved by introspection or corpus statistics alone.

1.1 Distributional vs. decompositional similarity

According to the distributional hypothesis, words that occur in similar contexts tend to have similar meanings (Firth 1957; Harris 1954; Miller & Charles 1991; Turney & Pantel 2010; Clark 2015). Since this hypothesis can be operationalized through approximation of word meaning "by the patterns of co-occurrence of words in corpora from statistical semantics" (Baroni et al. 2014: 241), the distributional hypothesis has become the main starting point for current research in computational semantics. In traditional count-based approaches to the distributional similarity of words, a word's meaning is typically a point in a high-dimensional vector space, where the dimensions of the vector correspond to context items, e.g. co-occurring words, and the coordinates of the vector are defined by the strength of these context items, e.g. co-occurrence counts. Contextual – or more precisely, distributional – similarity then becomes proximity of word meanings in the vector space. In this paper, I use distributional semantic models extracted from corpus data with neural network architectures that are referred to as "word embeddings" (Mikolov et al. 2013; Pennington et al. 2014). The relevant differences between word count models and word embeddings are that (i) at the quantitative level, word count models are high-dimensional while embedding models are low-dimensional and (ii) qualitatively, the dimensions of count models correspond to actual words, while the dimensions produced by embedding models can be thought of as soft clusters of context items that do not correspond to actual words (Levy & Goldberg 2014).

In a prototypical distributional model of word meaning, words that are most similar in meaning to the verb *laugh* are *funny*, *cry* and *tear*. The co-occurrence contexts of *laugh* thus reflect that a laughing event usu-

ally takes place in a funny situation, and often goes along with tears and crying. But the distributional similarity of words is not the only way in which words, and in particular verbs, can be judged to be similar. For example, according to the so-called unaccusative hypothesis (Perlmutter 1978), verbs like *laugh*,*work*, *sleep* and *run* are semantically similar in that they share the lexical entailment of an agentive meaning of their single argument – as opposed to verbs like *stumble*, *die*, *arrive* and *crystallize*, which are semantically similar because they entail a patient-like meaning of their argument.[1] But since verbs like *laugh*, *work*, *sleep* and *run* are highly unlikely to co-occur in similar contexts, unergative verbs cannot be semantically similar according to the distributional hypothesis.

In contrast to the distributional similarity of words, which can be read off the surface distribution of words in a corpus, the semantic similarity of intransitive verbs is semantically determined at a covert level of lexical representation and reflected by syntactic properties (Levin & Rappaport Hovav 1995) like e.g. auxiliary selection in the present perfect in German (Wunderlich 1985; Grewendorf 1989). Unergative verbs (1a) select HAVE in the present perfect, whereas unaccusative verbs (1b) select BE.

(1) a. *Maria hat gelacht.*
Maria HAVE laugh
'Maria has laughed.'
b. *Maria ist gestolpert.*
Maria BE stumble
'Maria has stumbled.'

[1]While there seem to be robust intuitions that there is a semantic difference between intransitive verbs that entail an agent-like meaning of their single argument and intransitive verbs that entail a patient-like meaning of their single argument, the grammatical realization of this intuitive dichotomy has been subject to debate, in particular with respect to the question for whether or not the unaccusative hypothesis pertains to a binary distinction or not, see e.g. Sorace (2000) for a gradient analysis of the unaccusative hypothesis based on data from auxiliary selection and Levin & Rappaport Hovav (1995) for a more general overview of the problem set. In this paper I make the more or less standard but simplifying assumption that the unaccusativity hypothesis makes a claim about a split in the syntax and semantics of intranstive verbs. Whether or not the methodology presented in this paper supports a specific view on the nature of the unaccusative hypothesis is a question I leave to further research.

Semantically, unaccusativity is determined by an intuition Dowty (1991: 605) characterizes as follows: "intransitive predicates argued to be unaccusative on syntactic grounds usually turned out to entail relatively patient-like meanings for their arguments [...], while those argued to be syntactically unergative were usually agentive in meaning." Levin & Rappaport Hovav (1995: 91) propose that unergative verbs describe internally caused events in which "inherent properties of the single argument like will, volition, emotion or physical characteristics are "responsible" for bringing about the eventuality" that the verb describes. Unaccusative verbs describe externally caused events for which an agent, an instrument, a natural force or a circumstance has "immediate control over bringing about the eventuality described by the verb" (Levin & Rappaport Hovav 1995: 92).

The semantic distinction between unergative verbs like *laugh* and unaccusative verbs like *arrive* is often represented through lexical decompositions as in (2) (see e.g. Levin & Rappaport Hovav 1995).

(2) a. laugh $\rightsquigarrow$ DO laugh
b. arrive $\rightsquigarrow$ BECOME arrived

In (2), the lexical meaning of the verbs *laugh* and *arrive* is decomposed into recurrent (word-overarching; blue) and idiosyncratic (word-specific; red) meaning elements.

Conceptual constants specify idiosyncratic properties of the event described by a verb. For example, *laugh* sets apart the manner of a laughing action from the manner of other actions, and *arrived* sets apart the state of having been arrived from other states. The unique vector representations assigned to words by vector space models of word meaning under the assumption of the distributional hypothesis capture exactly this kind of word-specific semantic information, given that what sets apart the manner of a laughing action from other actions are the specific contexts in which a specific a verb like *laugh* is used to describe the manner of a laughing event.

The predicates DO and BECOME are conceptual primitives that are not specific to a single verb, but recur in the decomposition of whole classes of verbs. DO is a meaning component that is present in the lexical decompo-

sition of all verbs that describe an action and entail an agentive meaning. BECOME is present in all verbs that describe a change of state and entail a patient-like meaning.

1.2 The open question for the components of word meaning

Against the background of the unaccusative hypothesis about intransitive verbs, the general problem with which the present paper is concerned is that work on the interpretation of distributional semantic representations like Levy & Goldberg (2014) fosters relatively clear-cut intuitions about how distributional semantic representations account for the meaning of specific words (and in particular nouns). But there are no similarly clear-cut intuitions about how distributional semantic representations of verb meaning may be understood to characterize concepts such as agency, volition or control, on the one hand and patiency or change of state, on the other. It thus remains an open question whether and how vector space models are also able to represent general concepts that are relevant to the lexical semantics of more than one word. A similar conclusion holds for theoretical approaches to verb meaning, since the type, number and determination of the recurring conceptual features relevant to verb meaning is a central, yet unsolved research problem. While theoretical work on lexical semantics assigns meaning components like agency or change of state a central role in the definition of thematic roles, it appears that "[t]here is perhaps no concept in modern syntactic and semantic theory which is so often involved in so wide a range of contexts, but on which there is so little agreement as to its nature and definition, as THEMATIC ROLE" (Dowty 1991: 547).

1.3 Methodological remarks

A combination of theoretical and computational approaches to lexical semantics poses at first a methodological challenge. Distributional semantics considers the main problem of lexical semantics to be a problem of empirical breadth: "[t]he problem of lexical semantics is primarily a problem of size: even considering the many subregularities found in the content lexicon, a hand-by-hand analysis is simply not feasible for the thousands of elements that populate the content word lexicon" (Baroni et al. 2014: 246). Accordingly, distributional semantic models are assessed ac-

cording to their extrinsic value, which can be measured by the performance of a certain model in a downstream task that evaluates a model's predictions against a large set of data annotated according to a gold standard. But the extrinsic assessment of computational approaches to semantics leaves open the question for whether and how computational models not only accomplish a certain task but also have an intrinsic explanatory value. As Lenci (2014) argues with a case study on the distributional classification of Italian verbs, the reproduction of a theoretically defined gold standard of verb classification does not indicate what the concepts or semantic features are like that underlie the classification. In contrast, theoretical approaches consider the main problem of lexical semantics to be a problem of analytical depth (Levin & Pinker 1991: 1):

> If *to paint* means 'cause to be covered with paint', why isn't it *painting* when a paint factory explodes or when Michelangelo dips his brush into the can [...]? These particular definitions can be patched up, but sceptics foresee a never-ending need for such patching, with no real increase in watertightness.

As a way out of this problem, theoretical lexical semantics proposes to focus on recurrent components in the decomposition of word meaning that can be used to define the meaning of verbs in terms of verb classes. But since there is no principled way to pin down the meaning components of a verb, the components of lexical decomposition have to be stipulated rather than being derived them from empirical observations, as Van der Leek (1996) argues with a case study on the conative alternation.

Given that theoretical and computational approaches to lexical semantics have clearly defined and widely agreed methodological standards, a middle ground between theoretical and computational semantics is likely to fall short of the established expectations of both theoretical and computational approaches to lexical semantics. The different focus of theoretical and distributional approaches to word meaning is easily mistaken as excluding an interoperability of the two methods, since bold and simple, from the viewpoint of theoretical lexical semantics, distributional approaches to word meaning fail to account for systematic, non-idiosyncratic aspects of word meaning, whereas from the viewpoint of distributional approaches to word meaning, theoretical lexical semantics fails to account

for idiosyncratic aspects of word meaning. The goal of the present paper is to widen the view on both theoretical and computational lexical semantics by bringing together the benefits of both approaches in the examination of the question whether, and if yes, how, general concepts like agency and patiency (represented by primitive predicates like DO and BECOME), respectively, are reflected in the embeddings of intransitive verbs.

1.4 Outline of the paper

I approach the problem of identifying recurrent meaning components in distributional semantic representations with the following steps.

First, the learning objective for word embeddings is distributional similarity, not the kind of decompositional similarity underlying the unaccusative hypothesis. I deal with this problem in §2 by fine-tuning general-purpose word embeddings of German and English intransitive verbs with the objective of learning to distinguish between unaccusative and unergative verbs. To identify the specific "surplus" that fine-tuning makes to the embeddings, I compute the difference vectors that represent the contribution of fine-tuning by subtracting task-specific intransitive verb embeddings from the corresponding general-purpose embeddings.

Second, since the difference vectors computed in §2, like general word embeddings, are dense and continuous, they are uninterpretable by humans. I address this problem by rendering the word embeddings interpretable through approximation of their meaning with their nearest neighbors in the embedding space (where the nearest neighbors of a given vector are the most proximate word embeddings in the embedding space).

Third, we do not know a priori how general concepts like those represented by DO and BECOME might be represented in the difference vectors. To account for this problem, in §4 I search the nearest neighbors of the difference vectors for meaningful and diagnostic linguistic cues that have been proposed in the theoretical literature to indicate an agentive or patient-like meaning. This qualitative inspection reveals that the nearest neighbors correspond to prototypical linguistic realizations of unergativity, like *-er* nominals and intentional actions for the unergative difference vector and descriptions of unintentional change of state events for the unaccusative difference vector. I conclude that the effect of retraining is neither trivial nor random, but instead captures surprisingly well the proper-

ties that linguists have found to be characteristic of agentive and patient-like verb meanings. In turn, the more general insight that I discuss in §5 is that the concepts that make up the meaning components of verbs are not directly encoded in embeddings by increasingly abstract terms as in ontologies like WordNet (Fellbaum 1998), but through patterns of word formation.

Finally, to further narrow down the concepts acquired through fine-tuning, in §6 I abstract away from specific difference vectors by estimating a linear regression model of the correlation between retrained and baseline unergative and unaccusative embeddings. The regression model allows to transfer the effect of retraining to other linguistic domains where the same distinction between DO and BECOME has been argued to be relevant: agentive *-er* and patient-like *-ion* nominals, respectively. The regression model also captures more complex conceptual generalizations, e.g. when the regression model for atelic DO and telic BECOME is used to transform the embeddings of mass and count nouns (following Bach (1986)). §7 concludes.

2 Data and methods

2.1 German and English word embeddings

The basis for the results reported in the present paper are German and English word embeddings learned with Word2Vec (Mikolov et al. 2013; SGNS: skip-gram with negative sampling, CBOW: continuous bag of words with hierarchical softmax, 300 dimensions, using 10w(ord) and 5w windows). The German embeddings were learned from SdeWac, a 0.88 billion word corpus of parsable German web data (Faaß & Eckart 2013), where the embedding space has a vocabulary of 237615 words. The English embeddings were learned from ukWac, a 1.3 billion word corpus of English web data (Ferraresi et al. 2008), where the embedding space has a vocabulary size of 70950 words. In the following, I refer to the German and English Word2Vec verb embeddings described in this section as "baseline embeddings". BERT embeddings (Peters et al. 2018) for German and English verbs were extracted from the multi-cased L-12/H-768/A-12 model,[2] using mean reduce as a pooling strategy on the last two output layers. In

[2]Downloaded from `https://github.com/google-research/bert`

the present paper I use BERT embeddings only for comparison for the following reasons. First, out-of-the-box BERT embeddings do not perform better in capturing unaccusativity than Word2Vec embeddings (see §2.4). Second, contextualized embeddings like BERT represent word meaning only relative to a context, which makes it difficult to pin down meaning components like DO and BECOME that words have independent of their context. Third, to rule out an interference of overt syntactic properties of unaccusativity like auxiliary selection during training of the embedding model, the Word2Vec embeddings used in this paper were won from corpus data filtered to consist only of content words. In comparison to BERT, Word2Vec is a computationally cheap way to learn embeddings and thus Word2Vec models can be trained from scratch without investing huge amounts of computation time.

2.2 Intransitive Verb Dataset

SdeWac was parsed with the syntactic and semantic dependency parser described in Björkelund et al. (2010). I extracted verbs that the parser saw more than 90 percent in an intransitive construction together with the semantic role label of the single argument (grammatical subject or grammatical object). I manually corrected the semantic role labels, using auxiliary selection in the present perfect (see (1)) as a diagnostic. In a further step of cleaning, I removed two classes of intransitive verbs that have been argued to involve an unaccusativity mismatch (Zaenen 1988) and thus are not unambiguously unergative or unaccusative, so-called verbs of emission and particle verbs of directed movement. In total, I ended up with a vocabulary of 972 unergative and 840 unaccusative German verb embeddings. Since English doesn't have reliable markers of unaccusativity (such as auxiliary selection), determining whether an English intransitive verb is unergative or unaccusative is more involved. I thus relied on existing lexical resources and used a subset of the unambiguously internally and externally caused verbs listed in the appendix of Levin & Rappaport Hovav (1995). As examples of externally caused verbs, I chose the classes of "alternating change of state verbs" and "cooking verbs" (251 verbs, class labels according to Levin 1993). For internally caused verbs, I chose "run verbs", verbs that partake in the unspecified object alternation and verbs that alternate with a cognate object construction (275 verbs). I used the

Embedding	linear SVM
baseline-SGNS-10w-de	fail to converge
retrained-SGNS-10w-de	1.0
baseline-CBOW-10w-en	0.89
retrained-CBOW-10w-en	0.99
BERT multi-cased-de	fail to converge
BERT multi-cased-en	fail to converge

Table 1 F1-score (10-fold cross-validation) of the binary classification of unergative/unaccusative verb embeddings with linear SVM.

classification of Levin & Rappaport Hovav (1995) to label intransitive verbs as subcategorizing either a grammatical subject or a grammatical object. The annotations for English nouns as "regular count nouns" and "regular mass nouns" were taken from the Bochum Countability Lexicon (Kiss et al. 2016).

2.3 Fine-tuning of verb embeddings through retraining

I fine-tuned the baseline embeddings for the German and English intransitive verbs through retraining the embeddings on a binary classification task on the unergative/unaccusative distinction, using RMSprop as the optimization algorithm and binary crossentropy as a loss function. For the retraining, I used a simple neural network architecture consisting of an embedding layer fully connected to a single output neuron. The resulting embedding layer is of size 1812 × 300 for German and 526 × 300 for English. The output neuron has a sigmoid activation function to yield a continuous probability distribution over the binary labels. To make sure the embeddings memorize the distinction between unergative and unaccusative verbs, the embeddings were overfit to 1.0 accuracy on the intransitive verb dataset.

2.4 Effect of fine-tuning: Linear separation

The effect of retraining is a linear separation of unergative and unaccusative verb embeddings, which becomes clearly visible in Figures 1 and 2, where the embeddings are projected down to two dimensions with PCA. In numbers, the linear separability achieved through retraining manifests itself in the improvement of the F1 score of a linear SVM classification task to

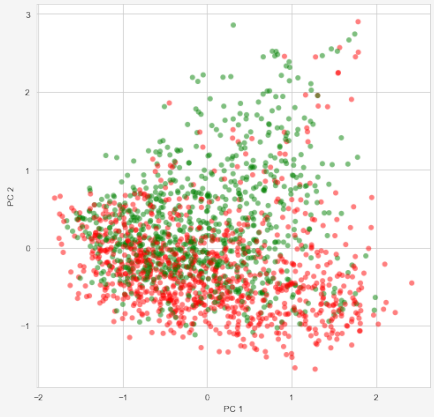

Figure 1 Baseline intransitive verb embedding space projected onto two dimensions with PCA (green = unaccusative verbs, red = unergative verbs)

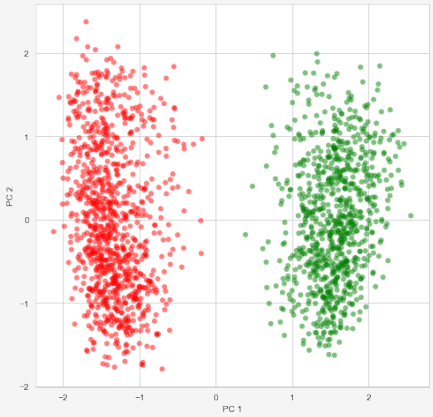

Figure 2 Retrained intransitive verb embedding space projected onto two dimensions with PCA.

near perfect, as reported in Table 1. The already good performance of linear SVM classification for the baseline English verb embeddings may be due to the selection of intransitive verbs, where I took great care to select only verbs that Levin & Rappaport Hovav (1995) classify unambiguously as unergative or unaccusative. It should be noted that state-of-the-art contextual embeddings like out-of-the-box BERT do not perform better than the classical Word2Vec embeddings, and thus that there is a genuine advantage in using contextualized embeddings over Word2Vec embeddings.

3 Quantitative evaluation

With respect to the quantitative effect of retraining, the first thing to rule out is that the weight updates through retraining are trivial. The weight updates would be trivial if the fine-tuned embeddings would have been changed in a way such that unergative and unaccusative embeddings are shifted to clearly distinct regions of the embedding space, but where the retrained embeddings would no longer be semantically similar to the baseline embeddings from which they are derived. To rule out that the weight updates are trivial in this sense, I consider the overlap in nearest neighbors between baseline and retrained embeddings. Here, and in what fol-

lows, I follow Levy & Goldberg (2014) and calculate the nearest neighbors of a word embedding with the dot-product, and correlate the magnitude of the dot-product with semantic similarity.

The overlap in the 10 nearest neighbors of the untrained and retrained verb embeddings is 76% for the German 10w SGNS model, and 97% for the English 10w CBOW model. These differences between German and English are likely to result from the difference in vocabulary size. Given that the German vocabulary is much bigger than the English vocabulary, we expect that for a given word, there is a high number of nearest neighbors with a small distance, and thus that even small changes in the embedding weights leads to a difference in the calculation of nearest neighbors. This explanation is supported by the fact that 96% of the top neighbors of the baseline German verb embeddings are also neighbors of the corresponding retrained verb embedding, whereas 87% of the top nearest neighbors of English baseline embeddings are neighbors of the corresponding retrained verb embeddings. The retrained verb embeddings thus basically retain the position of the baseline embedding in the embedding space, which suggests that retraining optimizes only a small subspace of the full embedding space without affecting the distributional similarities encoded in the embedding space as a whole. Injection of linguistic knowledge into word embeddings through retraining thus constitutes a cheap and effective way to improve the performance of word embeddings in classification problems that are difficult to approach on the basis of the distributional hypothesis alone, a finding the further exploration of which I leave to future research.

4 Qualitative interpretation

Given that the quantitative effect of retraining is non-trivial, the consequent question is whether the linear separation of unergative and unaccusative embeddings achieved through retraining is also systematic and non-random, i.e. whether the weights updates through retraining can be given a qualitative explanation that is related to the lexical decomposition of intransitive verbs. For this qualitative investigation of the retrained embeddings, I isolate the “surplus” that the retraining makes to the embeddings by subtracting for each of the embeddings of our German and English vocabulary the weights of the baseline embedding from the weights

of the retrained embedding as in (3) and then approximate the difference vectors with their nearest neighbors in the embedding space of the baseline embeddings.

(3) difference vector = retrained embedding − baseline embedding

One might wonder, as a reviewer does, about the mathematical soundness of applying operations of linear algebra on embedding spaces, given that embedding spaces like Word2Vec (and the retrained embeddings used in this paper) are learned through the application of non-linear operations like the sigmoid function. In fact, it is one of the surprising properties of word embedding models like Word2Vec that the model learned by the algorithm encodes certain semantic properties that can be revealed through the application of linear algebra to word embeddings. Most famously, word embedding models can be employed to solve word analogy tasks like "man is to woman like king is to ?" with methods of linear algebra. Linear algebra has also been used to model semantic aspects of morphological derivation (Padó et al. 2016) or verb formation (Pross et al. 2017). In a manner of speaking, the method employed in the present paper can thus be understood as an attempt to reveal with methods of linear algebra the properties of embedding spaces that correspond to the unaccusative hypothesis. The difference vectors of unergative and unaccusative embeddings represent the relevant semantic information that is responsible for the linear separability reported in Table 1. The first interesting observation to be made is that the approximation of the difference vectors with their nearest neighbors crystallizes in a small and restricted set of shared and recurrent nearest neighbors. For example, if we calculate the 10 nearest neighbors for each intransitive verb, there are 18120 possible nearest neighbors for the German verbs. But the approximation of the difference vectors in the 10w-SGNS German model results in a total of 10 shared and recurring nearest neighbors for the unergative difference vectors and 16 shared and recurring neighbors for the unaccusative difference vectors. The same observations holds for the English difference vectors. Out of a total of 5260 possible 10 nearest neighbors of the English verbs, in the 10w-CBOW English model the difference vectors for the unergative embeddings consist of 15 recurring neighbors and and 24 neighbors for the unaccusative difference

vectors. In the next section, I consider the nearest neighbor approximation of difference vectors in more detail.

4.1 Unergative difference vectors

Consider first the approximated German and English difference vectors for German 10w-SGNS embeddings and English 10w-CBOW embeddings in (4) and (5), where I list the six most frequent nearest neighbors of the unergative and unaccusative difference vectors. I cleared the neighbor lists of proper names and derivationally related words. The chars following the neighbor represent the POS-Tag (n=noun, v=verb, a=adjective) and the numbers following the POS-Tag indicate how often that neighbor occurred in the approximations of the difference vectors of the set of verbs in question.

(4) *Prüferinnen.n.972 Informatikkaufmann.n.972*
'examiners' 'IT merchant'
Gruppenarbeitsraum.n.972 Diätassistent.n.972
'group office' 'diet assistant'
Kulturmanager.n.952
'culture manager'

(5) beginner.n.251 beginners.n.251 sewing.n.251 salary.n.251
crafty.a.251 ceilidh.n.246.0

An obvious cue of an agentive meaning in the neighbor approximations of the unergative difference vectors are *-er* nominals like *Prüfer* ('examiner'), *beginner* and job descriptions like *Kaufmann* ('merchant') or *Assistentin* ('assistant') , but also descriptions of agentive actions like *sewing* and related adverbials like *crafty*. The general tendency of an agentive meaning that appears in the approximated unergative difference vectors is stable across the statistical variation to retraining introduced by optimization with RMSprop. (6) and (7) are the six overall most frequent neighbors of the difference vectors of 50 runs of retraining the German 10w-SGNS model and the English 10w-CBOW model, respectively. Again, we encounter clear indicators of agentive meaning, job descriptions like *Redakteur* ('editor') or *illustrator*, agentive verbs like *rülpsen* ('burp') or agentive nouns like *sleeps*.

(6) *Transgender.n.1976 Redakteur.n.1944 Chef.n.1944*
'transgender' 'editor' 'chief'
Familiensonntag.n.1944 rülpsen.v.1942 Akustiker.n.1650
'family sunday' 'burp' acoustician'

(7) prance.v.3491 toy.n.1408 groovy.n.1384 illustrator.n.1296 sleeps.n.1170 replies.n.1006

Indication of an agentive meaning is also stable across window size. (8) and (9) are the approximations of the unergative difference vectors using five word window SGNS embeddings for German and 5w CBOW embeddings for English. Indicators of agentive meaning are again *-er* nominals like *Comiczeichner* ('cartoonist') or *tinker* and job descriptions like *illustrator*, but also intentional verbs like *freuen* ('be glad') and mental adjectives like *stinksauer* ('spitting mad').

(8) *Comiczeichner.n.972 freuen.v.972 klingeltöne.n.972*
'cartoonist' 'be glad' 'ring tones'
Mitgliedermagazin.n.972 rülpsen.v.971 stinksauer.a.677
'member magazine' 'burp' 'spitting mad'

(9) warranties.n.251 tinker.n.251 cordless.n.251 illustrator.n.251 difficulties.n.250 enquiries.n.250

4.2 Unaccusative difference vectors

In contrast to the indicators of agentive meaning that are present in the approximations of the unergative difference vectors, the unaccusative difference vectors for the German 10w-CBOW and English 10w-SGNS model in (10) and (11) are characterized by verbs that describe non-intentional change-of-state events like *ionisieren* ('ionize'), *purify* or *refract*, *-ion* nominals like *separation*, *eruption*, inanimate yet causally active substances like *toxin* and adjectival nominalizations such as *redness*.

(10) *Kontinentalplatte.n.840 ionisieren.v.840.0 Bremsvorgang.n.840*
'continental plate' 'ionize' 'braking process'
Vorderreifen.n.840 aushärten.v.839 Ladungstrennung.n.825
'front tire' 'harden' 'charge separation'

(11) purify.v.274 refract.v.265 redness.n.265 adverse.a.135 toxin.n.37

eruption.n.20

As for the unergative difference vectors, indicators of patient-like meanings in the unaccusative difference vectors are stable across window sizes. (12) and (13) are the nearest neighbors of the unaccusative difference vectors for five word window SGNS (German) and CBOW (English) embeddings. Again, we encounter *-ion* nominals like *reaction*, *circulation* and nominals that describe non-intentional change-of-state events like *heat* as well as inaminate yet causally active substances (*oxygen*, *metastasis*) and their properties (*flammable*).

(12) *Kernreaktion.n.840 Tochtergeschwülste.n.840*
'nuclear reaction' 'metastasis'
Unfallereignis.n.840 Verwitterung.n.840 Katarakt.n.840
'event of accident' 'weathering' 'cataract'
Ausschwemmung.n.836
'washout'

(13) heat.n.275 circulation.n.275 oxygen.n.275 flammable.a.275
starch.n.274 inefficient.a 268

The general tendency of a patient-like meaning that appears in the approximated unaccusative difference vectors is also stable across the statistical variation to retraining introduced by RMSprop. (14) and (15) are the six overall most frequent neighbors of the difference vectors of 50 runs of retraining the German 10w-SGNS embeddings and the English 10w-CBOW embeddings, respectively. Again, we see clear indicators of a non-intentional change-of-state meaning, like *Dickenwachstum* ('growth in girth') or *warming*.

(14) *einspritzen.v.2523 Dickenwachstum.n.2431 Zugspannung.n.1682*
'inject' 'growth in girth' 'tension'
Gerinnsel.n.1682 Verbuschung.n.1682 Schwerefeld.n.1682
'clot' 'scrub encroachment' 'gravitation field'

(15) moisture.n.144 temperature.n.114
warming.n.55 unenforceable.a.52 cremation.n.50 detoxify.v.40

In sum, an informed linguist is able to interpret the nearest neighbors of

the unergative and unaccusative vectors as characterizations of the distinction between agent- and patient-like meanings by detecting word formation patterns connected to unergativity, such as *-er* nominals, and shared lexical entailments of proto-agent and proto-patient properties in the sense of Dowty (1991). Since these observations can be obtained independently for both English and German intransitive verbs, this suggests that the representations of the semantic correlates of the unaccusative hypothesis by the approximated difference vectors of intransitive verbs are not random outliers but rather point towards a systematic effect of retraining of intransitive word embeddings. One explanation for this systematic effect may be that retraining of the embeddings is a method for strengthening those latent dimensions of the embedding space that involve the same recurrent meanings relative to the retraining objective. In the next §, I consider these latent dimensions of the embedding space in more detail.

5 Word formation and concept representation

In §4, I demonstrated that the quantitative effect of retraining qualitatively amounts to the addition of non-trivial and non-random information to the baseline embeddings. Inspection of the approximated difference vectors of unergative and unaccusative embeddings furthermore suggested that this information does not correspond to the meaning of single words but instead is related to general concepts like agency and intentionality for the unergative difference vectors and non-intentional change of state for the unaccusative difference vectors. An eye-catching observation in this respect is that these concepts are represented in the nearest neighbors of the difference vectors through systematic patterns of word formation. One particularly telling example are nominals derived with *-er*/*-or* for the unergative difference vectors and *-ion*/*-ung* nominals for the unaccusative difference vectors. The observation that word formation, and in particular the semantic distinction expressed by English nominals derived through suffixation with *-er*/*-ion* is highly diagnostic is reinforced by the overall proportion of these nominals in the nearest neighbor approximation of the difference vectors. For example, in the difference vectors of 50 runs of retraining the English 10w CBOW intransitive verb embeddings, there are 443 *-er* nominals in the unergative difference vectors, but only two *-ion* nominals; and conversely, there are 458 *-ion* nominals in

the unaccusative difference vectors, but only 17 *-er* nominals. One way to explain this observation is that the appearance of systematic patterns of word formation is no coincidence, since derivational morphology is one way in which languages overtly mark word-overarching meaning components. In fact, the word formation patterns of *-er*/*-ion* nominals have been argued in the theoretical literature to correlate with the distinction between an agentive and a patient-like meaning. As seen in (16), *-er* nominalizations have an agentive semantics and can only be derived from DO-verbs (Levin & Rappaport 1988).

(16) a. dancer, worker, dreamer
b. *faller, *asleeper, *arriver

In contrast, *-ion* nominalizations as in (17) have a passive semantics and can only be derived from (transitive) BECOME-verbs (Alexiadou 2001):

(17) a. destruction, explosion

These considerations suggest that the concepts that make up recurrent meaning components are not directly encoded in embeddings by increasingly abstract terms as in ontologies like WordNet (Fellbaum 1998), but through patterns of word formation. If this hypothesis is on the right track, we expect to see the same effect in other domains where patterns of word formation are semantically indicative. This expectation is borne out for the case of mass/count nouns. (18) and (19) present the 10 most frequent recurring neighbors of the difference vectors for English word embeddings of 1270 regular count and 2233 mass nouns. The neighbors in (18) and (19) were calculated using the same pipeline as for the unergative and unaccusative verbs, consisting of retraining noun embeddings with a binary classification task (on the mass/count distinction), computation of difference vectors and approximation with their nearest neighbors. As for the unergative/unaccusative verbs, the nearest neighbors of mass/count difference vectors consist of a small and restricted set of recurrent elements. The approximation of the 10w-CBOW difference vectors of English mass nouns consists of 19 recurrent nearest neighbors out of 22330 total possible nearest neighbors, and the approximation of the 10w-CBOW difference vectors of English count nouns consists of 35 recurrent neighbors out

of a total of 12700 possible neighbors.

(18) *10 most frequent neighbors of the difference vectors of English mass nouns*
tannin.n.1270 absorption.n.1269 sensuality.n.1269 warmth.n.1260 tightness.n.1256 metabolism.n.1235 cheques.n.1025 acidity.n.900 solubility.n.836 irritability.n.726

(19) *10 most frequent neighbors of the difference vectors of English count nouns*
keypad.n.2025 aisle.n.2025 doorway.n.2025 door.n.2024 verifiers.n.2019 policeman.n.2013 impale.v.2004 lock.v.1931 baccalaureate.n.746 row.n.520

Notably, in (18) we again encounter word-formation patterns as being indicative of the concepts of a mass noun, here in the form of *-ity* and *-ness* nominalizations. These nominalization formation patterns have been argued in the theoretical literature to derive abstract properties, "qualities" which are ontologically similar to mass noun (Francez & Koontz-Garboden 2017). The concept of a count noun, in contrast, seems to be encoded by the simple fact that the nearest neighbors in (19) – in contrast to the neighbors in (18) – can be counted. I take these observations to provide additional evidence for the claim that retraining adds non-trivial and non-random conceptual information to general-purpose word embeddings.

6 Lexical decomposition with distributional semantics

In §5, I showed through qualitative inspection of difference vectors that the surplus that retraining makes to word embeddings of verbs and nouns is neither trivial nor random, but instead captures the relevant high-level conceptual distinction on which the embeddings were fine-tuned. The observations I made, however, pertained to difference vectors of specific verbs and nouns. To represent the effect of retraining in a more general way independent of specific words, in this section I approximate the relation between the baseline and retrained intransitive English verb embeddings with a linear regression model that captures the effect of retraining as a linear transformation on embeddings. Casting the effect of retraining in such a general and abstract way allows me to investigate in more detail

the kind of concept acquired through retraining by applying the regression model to word embeddings that have been argued in the theoretical literature to involve the same kind of meaning components as intransitive verbs.

6.1 The regression model

To capture the effect of retraining, I estimated two regression models for the relation between (i) the unergative baseline and retrained embeddings and (ii) the unaccusative baseline and retrained embeddings. In (20), B is a coefficient matrix (more precisely, a diagonal matrix with the regression weights as diagonal elements) and Err an error term. Verb embeddings $x \in X_{baseline}$ are the independent variables of the model, and verb embeddings $y \in Y_{retrained}$ the dependent variables. I estimated the coefficient matrices and error terms of the linear regression equation with partial least squares regression on matrices X and Y, using the number of embedding dimensions as latent variables. Since I want to capture the effect of retraining, I overfit the estimated linear models for unergative and unaccusative verbs to an R^2 score of 1.0.

(20) $$Y_{retrained} = X_{baseline} * B + Err$$

I propose to use the estimated regression model for the relation between unergative and unaccusative baselines and retrained verb embeddings as an interpretation of the lexical decomposition of unergative and unaccusative intransitive verbs. (21) and (22) illustrate lexical decomposition with distributional semantics with the examples of unergative *laugh* and unaccusative *arrive*.

(21) $$\text{to laugh} \rightsquigarrow \text{DO } laugh \rightsquigarrow B_{do} * laugh + Err_{do}$$

(22) $$\text{to arrive} \rightsquigarrow \text{BECOME } arrived \rightsquigarrow B_{become} * arrive + Err_{become}$$

In (21) and (22), the conceptual primitives DO and BECOME are interpreted as the coefficients of the estimated linear regression model. As discussed in §1, word embeddings excel in capturing idiosyncratic aspects of word meaning. Thus, I propose that the interpretation of the idiosyncratic constants *laugh* and *arrived* is provided by the corresponding verb embeddings of *laugh* and *arrive*.

6.2 Evaluation

I evaluate the linear regression model by applying the linear transformation described by the model to linguistic domains that have been argued in the theoretical literature to involve the same kind of conceptual distinction. In the first evaluation, I apply the regression model for DO and BECOME to *-er*/*-ion* nominals, and in the second evaluation to mass and count nouns.

To evaluate the regression model against *-er*/*-ion* nominals, I collected all the 3812 "agentive" *-er* and 2357 "patient-like" *-ion* nominalization types contained in ukWac. Linear SVM classification of the English 10w-CBOW baseline embeddings for these nominals fails to converge. I then applied the linear regression model to the *-er*/*-ion* nominal baseline embeddings, as illustrated for the nominals *beginner* and *eruption* in (23a) and (23b), respectively.

(23) a. beginner $\rightsquigarrow B_{do} * \textit{beginner} + Err_{do}$
b. eruption $\rightsquigarrow B_{become} * \textit{eruption} + Err_{become}$

When the regression model for DO is used to transform the embeddings of *-er* nominals and the regression model of BECOME to transform the embeddings of *-ion* nominals as in (23), this improves the F1 score of the transformed English noun embeddings in a linear SVM classification task on the *-er*/*-ion* distinction to 0.97.

In the second evaluation, I test my regression model with a more far-fetched conceptual correlation than the distinction between agentive and patient-like nominals and agentive and patient-like verbs. Bach (1986) proposed a correlation between the concepts underlying the distinction between mass and count nouns, on the one hand, and DO and BECOME verbs on the other: mass is to atelic verbs (i.e. unergative verbs involving DO, see Zaenen (1988)) what count is to telic verbs (i.e. unaccusative and transitive verbs involving BECOME) as in (24).

(24) *atelic* :: *mass* = *telic* :: *count*

Linear SVM classification on the 10w-CBOW English baseline embeddings of the 3000 count and mass nouns from §5 fails to converge. But application of the DO regression model to the mass noun embeddings and the

BECOME regression model to the count nouns yields an F1 score of 0.97 for linear SVM classification of the transformed embeddings. To ensure this is not a random effect, I did a countercheck with interchanged regression models. When the DO-model is applied to count nouns and the BECOME-model to mass nouns, linear SVM classification of the transformed embeddings still fails to converge.

7 Conclusion and outlook

The main goal of the present paper was to bring together linguistic intuitions about word meaning with the empirical groundedness of computational approaches to lexical semantics derived from large corpora.

Before I summarize the findings of this paper and put into them into the context of larger research goals, I briefly return to the methodological issues of approaching questions of lexical semantics with methods of machine learning I already addressed in §1.3. A reviewer notes that the interpretation of word embeddings through approximation with their nearest neighbors, and in particular the interpretation of nearest neighbors as providing support of a certain hypothesis, makes the analysis "merely impressionistic", where one just uses one's own intuition to highlight neighbors that confirm the hypothesis, whereas neighbors that have nothing to do with unaccusativity are dismissed. I believe this is an important point because it pertains to the very question of linguistic theorizing. Broadly speaking, the word embeddings discussed in this paper can be understood as compressed representations of word usage derived from an empirical basis the size of which is beyond the limits of human capacity. Traditional linguistic theory is standardly based on a small set of carefully crafted examples. But exceptions to the rule of course also occur when theorizing about carefully selected examples, and the unaccusative hypothesis is one such topic in modern linguistics where there are countless exceptions to the rule, to the extent that a unified theory of intransitive verbs seems to become impossible (see Alexiadou et al. 2004 for an overview of the current state of the debate). The "impressionistic" method of interpreting approximated word embeddings through the eyes of an informed linguist may thus be considered as reproducing exactly the difficulties of theorizing about linguistic observations, in that they provide an unbiased, yet empirically grounded view of the complexities of linguistic

reality. Since it is difficult to determine which method of picturing natural language is the right one, instead of contrasting the methodological standards of the two approaches, I propose that striving for a combination of insights gained with methods of traditional linguistic theory and methods of machine learning may provide a more accurate picture of the inner workings of natural language,even if the picture may seem fuzzy and ad-hoc. This being said, the purpose of this paper is not to prove a certain point but rather to make an attempt at investigating how word embeddings of a certain data set relate to linguistic theorizing about the same data set.

From a theoretical point of view, the results obtained in this paper suggest that vector models of word meaning may in fact capture the high-level conceptual distinction related to the conceptual primitives DO and BECOME. The possibility to transfer the information gained through retraining of intransitive-verb embeddings to other domains shows that the same meaning component may in fact be present in intransitive verbs, derived *-er*/*-ion* nominals and mass and count nouns. Since word embeddings are learned from data independent of theoretical bias and intuitions, the case study presented in this paper suggests that word embeddings can be understood as providing a fresh perspective on the conceptual foundations of lexical semantics. The consequent challenge for the theoretical linguist is to make use of this new perspective on lexical meaning, since the observations made in the present paper to spell out in more detail the ontology of the parallels between the verbal and nominal domain that in the tradition of Bach have mainly been assessed in terms of a structural algebra. I leave the question for a unified ontology of verbs and nouns suggested by the case study in this paper as a challenge to future research.

From a computational point of view, it is important to note that I did not feed the relation between DO and BECOME and derived *-er*/*-ion* nominals or mass and count nouns into the model. The regression model thus captures general aspects of the linguistic intuition underlying the conceptual distinction between DO and BECOME. A possible explanation for this observation is that retraining strengthens latent semantic features of the baseline word embeddings. In the approach to semantic similarity based on the dot-product, feature strengthening finds expression in the magni-

tude of a vector. In fact, the regression weight matrices of the regression models for DO and BECOME are not just diagonal matrices. They are near-perfect scalar matrices, where the mean of the DO diagonal vector is 0.84 (with a variance of 0.00091) and the mean of the BECOME diagonal vector is 0.936 (with a variance of 0.00037). One main effect of the application of the regression model is thus a stretching of embeddings by a certain scalar factor. From this point of view, the results presented in §6 suggest that the scalar factors we estimated with the regression model are indeed semantically meaningful. In turn, one may speculate whether general concepts are represented in vector space models of meaning like those that can be derived with the Word2Vec algorithm as scalar operations on embeddings. I leave a further exploration of this idea to future research.

The more general hypothesis that arises from the discussion in the present paper concerns the nature of embedding spaces by themselves. Embeddings of specific words only occupy a small part of the full (continuous and dense) embedding space. This paper showed that some of the vectors and operations in the embedding space that do not correspond to words are actually meaningful: they represent abstract concepts like those associated with DO and BECOME. The more general goal of future research that this paper envisages is to use linguistic insights as a roadmap of the terra incognita of embedding spaces in which meaningful vectors live that do not correspond to specific words. Ever since Fillmore (1968), who characterized the covert concepts encoded by verbs as "a set of universal, presumably innate, concepts which identify certain types of judgments human beings are capable of making about the events that are going about around them, judgments about such matters as who did it, who it happened to, and what got changed" (Fillmore 1968: 45), linguists have been working on developing frameworks in which concepts that refer to meanings that are not associated with words. Linguistic theory thus plays a central role in the identification of meaningful non-word vectors in embedding spaces by providing frameworks in which non-word concepts can be systematically investigated. In turn, bringing together the compact and efficient representations that have been developed in theoretical lexical semantics with the empirical grounding of computational approaches to word meaning in the spirit of the distributional hypothesis may provide a way out of the ever-growing resource demands of modern natural lan-

guage processing.

Acknowledgments I would like to thank the reviewers and participants of CSSP 2019 for valuable feedback on earlier versions of this paper. The research reported in this paper has been supported by DFG grant PR1860/1.

References

Alexiadou, Artemis. 2001. *Functional structure in nominals – nominalization and ergativity*. Amsterdam: John Benjamins.

Alexiadou, Artemis, Elena Anagnostopoulou & Martin Everaert (eds.). 2004. *The unaccusativity puzzle: Explorations of the syntax-lexicon interface*. Cambridge University Press.

Asher, Nicholas, Tim Van de Cruys, Antoine Bride & Márta Abrusán. 2016. Integrating type theory and distributional semantics: A case study on adjective–noun compositions. *Computational Linguistics* 42(4). 703–725.

Bach, Emmon. 1986. The algebra of events. *Linguistics and Philosophy* 9. 5–16.

Baroni, Marco, Raffaela Bernardi & Roberto Zamparelli. 2014. Frege in space: A program of compositional distributional semantics. *LiLT* 9. 241–346.

Björkelund, Anders, Bernd Bohnet, Love Hafdell & Pierre Nugues. 2010. A high-performance syntactic and semantic dependency parser. In *Coling 2010: Demonstrations*. 33–36. Coling 2010 Organizing Committee. `http://aclweb.org/anthology/C10-3009`.

Boleda, Gemma. 2020. Distributional semantics and linguistic theory. *Annual Review of Linguistics* 6. 213–234.

Clark, Stephen. 2015. Vector space models of lexical meaning. In Shalom Lappin & Chris Fox (eds.), *The handbook of contemporary semantic theory*, 493–522. Wiley Blackwell 2nd edn.

Dowty, David R. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Faaß, Gertrud & Kerstin Eckart. 2013. SdeWaC – a corpus of parsable sentences from the web. In Iryna Gurevych, Chris Biemann & Torsten Zesch (eds.), *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology*.

Fellbaum, Christiane. 1998. *Wordnet: An electronic lexical database*. Cambridge, MA: Bradford Books.

Ferraresi, Adriano, Eros Zanchetta, Marco Baroni & Silvia Bernardini. 2008. Introducing and evaluating ukWaC, a very large web-derived corpus of English. In *Proceedings of the 4th Web as Corpus Workshop (WAC-4)*. 47–54.

Fillmore, Charles J. 1968. The case for case. In Emmon Bach & Robert T. Harms (eds.), *Universals in linguistic theory*, 1–90. Holt, Rinehart, and Winston.

Firth, John R. 1957. *Papers in linguistics, 1934–1951*. Oxford: Oxford University Press.

Francez, Itamar & Andrew Koontz-Garboden. 2017. *Semantics and morphosyntactic variation*. Oxford: Oxford University Press.

Grewendorf, Günther. 1989. *Ergativity in German*. Dordrecht: Foris.

Harris, Zellig. 1954. Distributional structure. In *Papers in structural and transformational linguistics*, vol. 10, 1456–1462. Dordrecht: Springer. doi:10.1007/978-94-017-6059-1_36.

Kiss, Tibor, Francis Jeffry Pelletier, Halima Husic, Roman Nino Simunic & Johanna Marie Poppek. 2016. A sense-based lexicon of count and mass expressions: The Bochum English countability lexicon. In *Proceedings of LREC*. 2810–2814. Portorož, Slovenia: European Language Resources Association (ELRA).

Lenci, Alessandro. 2014. Carving verb classes from corpora. In Raffaele Simone & Francesca Masini (eds.), *Word classes: Nature, typology and representations*, 17–36. John Benjamins.

Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.

Levin, Beth & Steven Pinker. 1991. Introduction. In *Lexical & conceptual semantics*, 1–8. Blackwell.

Levin, Beth & Malka Rappaport. 1988. Non-event *-er* nominals: A probe into argument structure. *Linguistics* 26. 1067–1083.

Levin, Beth & Malka Rappaport Hovav. 1995. *Unaccusativity at the syntax-semantics interface*. Cambridge, MA: The MIT Press.

Levy, Omer & Yoav Goldberg. 2014. Neural word embedding as implicit matrix factorization. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence & K. Q. Weinberger (eds.), *Advances in neural information processing systems 27*, 2177–2185. Red Hook, NY: Curran Associates, Inc.

McNally, Louise & Gemma Boleda. 2017. Conceptual vs. referential affordance in concept composition. In James A. Hampton & Yoad Winter (eds.), *Compositionality and concepts in linguistics and psychology*, 245–267. Cham: Springer International Publishing.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado & Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of NIPS*. 3111–3119.

Miller, George A. & Walter G. Charles. 1991. Contextual correlates of semantic

similarity. *Language and Cognitive Processes* 6. 1–28.

Padó, Sebastian, Aurélie Herbelot, Max Kisselew & Jan Šnajder. 2016. Predictability of distributional semantics in derivational word formation. In *Proceedings of coling*. 1285–1296. Osaka, Japan. `http://www.aclweb.org/anthology/C/C16/C16-1122.pdf`.

Pennington, Jeffrey, Richard Socher & Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of EMNLP*. 1532–1543.

Perlmutter, David M. 1978. Impersonal passives and the unaccusative hypothesis. In *Proceedings of the 4th Annual Meeting of the Berkeley Linguistics Society*. 157–190. Berkeley Linguistics Society.

Peters, Matthew E., Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee & Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of NAACL*. 2227–2237.

Pross, Tillmann, Antje Roßdeutscher, Sebastian Padó, Gabriella Lapesa & Max Kisselew. 2017. Integrating lexical-conceptual and distributional semantics: A case report. In Alexandre Cremers, Thom van Gessel & Floris Roelofsen (eds.), *Proceedings of the 21st Amsterdam Colloquium*. 75–85.

Sorace, Antonella. 2000. Gradients in auxiliary selection with intransitive verbs. *Language* 76. 859–890.

Turney, Peter D. & Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37. 141–188.

Van der Leek, Frederike. 1996. The English conative construction: A compositional account. In *Proceedings of the Chicago Linguistic Society 32*, vol. 32. 363–378. Chicago Linguistic Society.

Wunderlich, Dieter. 1985. Über die Argumente des Verbs. *Linguistische Berichte* 97. 183–227.

Zaenen, Annie. 1988. Unaccusatives in Dutch and the syntax-semantics interface. CSLI Report 123 Center for the Study of Language and Information Stanford, CA.