Distributionelle Repräsentationen als konzeptuelle Grundlage einer linguistisch fundierten Theorie der Semantik von Wortwurzeln

From distributions to roots – Towards a linguistically grounded theory of the conceptual underpinnings of verb meaning

Tillmann Pross, Stuttgart

Zusammenfassung

Eine etablierte Methode der lexikalischen Semantik ist die Unterscheidung von Verbklassen anhand konzeptueller Eigenschaften, z.B. indem man Verben die eine gerichtete Bewegung beschreiben von Verben unterscheidet die einen Zustandswechsel beschreiben. Die Art, Anzahl und Bestimmung der relevanten konzeptuellen Eigenschaften ist allerdings eine offene Forschungsfrage sowohl in theoretisch orientierten Ansätzen zur lexikalischen Semantik als auch in der Computerlinguistik. Das Projekt untersucht am Beispiel von intransitiven Verben wie und ob konzeptuelle Strukturen die aus distributionellen semantischen Repräsentationen abgeleitet sind eine neue Perspektive auf die in der lexikalischen Semantik gängigen Annahmen über die konzeptuellen Grundlagen der Analyse von Verben erlaubt. Das Ziel des Projekts ist die Entwicklung einer Proto-Theorie der konzeptuellen Interpretation von distributionellen semantischen Repräsentationen, die als Grundlage einer theoretisch adäquaten Analyse von Verbbedeutung dienen kann. Vereinfacht gesagt, untersucht das Projekt ob es einen systematischen, theoretisch fundierten und komputationell unterstützten Weg gibt, die drastische Dimensionsreduktion durchzuführen, die nötig ist um hoch-dimensionale distributionelle semantische Repräsentationen von Verben in solch niedrig-dimensionale konzeptuelle Repräsentationen zu überführen, wie sie in der lexikalischen Semantik verwendet werden. Auf einer übergreifenden Ebene zielt das Projekt darauf ab, eine Kombination von theoretischen und komputationellen Ansätzen für die Analyse von Verbbedeutung als einen möglichen Weg zu einer empirisch und theoretisch fundierten Analyse von Verbbedeutung zu motivieren.

Summary

A standard method in lexical semantics is to distinguish verb classes conceptually, e.g. by distinguishing verbs that describe a directed motion from verbs that describe a change of state. But the type, number and determination of the conceptual features relevant to verb meaning is an open research question in both theoretical and computational approaches to verb meaning. The project investigates, for the case of intransitive verbs, whether and how conceptual structures derived from distributional semantic representations of verbs provide a novel perspective on those conceptual structures that are standardly invoked in lexical semantics. The goal of the project is come up with a proto-theory of the conceptual interpretation of distributional semantic representations that can be input to theoretically inspired analyses of verb meaning. Figuratively speaking, the project investigates whether there is a systematic, theoretically informed and computationally supported way to perform the drastic dimension reduction that is required to convert a high-dimensional distributional semantic representation of a verb into a low-dimensional conceptual structure that can be understood as a theoretically reasonable and sensible re-

presentation of the conceptual meaning of that verb. On an overarching level, the project aims at showing that a combination of theoretical, lexical-conceptual and computational, usage-based approaches to verb meaning may pave the way towards an empirically grounded and theoretically sound theory of verb meaning in its entirety.

Project Description

From distributions to roots – Towards a linguistically grounded theory of the conceptual underpinnings of verb meaning

Tillmann Pross, Stuttgart

1 State of the art and preliminary work

1.1 State of the art

A theory of lexical representation is key to compositional theories of the meaning of phrases and sentences. One of the main challenges in lexical semantics is that general theories of how systematic aspects of word meaning are represented must be induced from research on how idiosyncratic aspects of particular word meaning are represented. This challenge has been met quite differently in lexical-conceptual and distributional approaches to word meaning.

Lexical-conceptual semantics In theoretical linguistics, a widely adopted hypothesis that drives research in lexical semantics is that "syntactic properties of phrases reflect, in large part, the meanings of the words that head them" (Levin and Pinker, 1991, p. 3). One way to represent these syntactically relevant components of meaning is to decompose a verb's meaning into a fixed set of primitive predicates (like 'become') and constants from a limited set of semantic types (like 'broken'). For example, (1-b) is the lexical decomposition of (1-a).

(1) a. The window is broken.
b. [*y become BROKEN*]

Typically, verbs of the same semantic class have common substructures in their decompositions. E.g. all verbs of change of state involve a substructure with the primitive 'become', in which a constant names the state filling the second argument of 'become'. Such general templates are what Pinker (2013) calls "thematic cores" and Levin and Rappaport Hovav (1995) refer to as "lexical semantic templates". But syntactic properties of phrases have been argued to reflect even more fine-grained distinctions among verbs. For example, to explain the grammaticality of verbs in the conative construction, i.e. She cut at the bread vs. *She broke at the bread, Guerssel et al. (1985) proposed that the relevant distinction is of a conceptual nature. In the terminology of Pinker (2013), the relevant distinction is realized by a "narrow-range" lexical rule: cut is a verb of motion, contact and causation whereas break is a verb of pure causation. Consequently, the concepts of motion, contact and causation must be represented in the particular meaning of a verb in a way that syn-

tax can be sensitive to. That is, syntax not only provides clues to the general "templatic" aspects of verb meaning but also to narrow-range constraints on the usage of a particular verb. As e.g. Levin (1993) shows impressively, when we extend the search for such syntactically represented conceptual distinctions to a wider range of verbs and constructions, a systematic and fine-grained lexical-conceptual classification of "semantically cohesive" verb classes can be induced. I refer to this particular alternation-based approach of verb meaning in the following as the lexical-conceptual structure (LCS) approach to verb meaning. It should be noted, however, that the general idea that lexical entries involve both a templatic structure and a conceptual specification is not specific to the LCS framework but is also assumed in other theories of lexical semantics like "Semantic Forms" (Bierwisch, 2007), albeit motivated there on different grounds. In what follows, I understand the term 'conceptual meaning' to refer to meanings that are not associated with words, that is, with specific linguistic expressions belonging to certain syntactic categories. The same concept can be realized by different words. For example, the concept of 'directed manner of motion' can be expressed by the verb to walk or the noun a walk. Accordingly, I understand the terms 'word meaning' and 'lexical meaning' to refer to a meaning that is associated with a specific linguistic expression of a certain syntactic category, i.e. a word. That is, I wish to distinguish between those "connections we make between linguistic expressions and our conceptual structure, on the one hand, and the world, on the other" (McNally and Boleda, 2017, p. 251).

Distributional Semantics A popular computational approach to lexical semantics, namely distributional semantic models (DSMs), starts from the hypothesis that "words that occur in similar contexts tend to have similar meanings", see Turney and Pantel (2010) for an overview. Accordingly, the distribution of a word's contexts are considered central to the construction of a suitable meaning representation of that word. A DSM representation of the meaning of a word 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 similarity then becomes proximity of word meanings in the vector space. The DSM approach to word meaning is often illustrated by appeal to intuitions like the following (see e.g. Clark (2015)): football is similar in meaning to soccer since many of the words surrounding instances of football — within a contextual window of a sentence — are the same as the words surrounding instances of soccer. Theories of verb meaning like the LCS framework have been related to DSM approaches of word meaning with so-called "structured" DSM models (Baroni and Lenci, 2010), where DSM representations are not harvested out of an unstructured window of tokens surrounding a given word, but from the distribution of words in specific syntactic-semantic frames. When the semantic feature spaces of structured DSM representations of contextual similarity are input to supervised classification or unsupervised clustering algorithms, verb classes

similar to those identified in the LCS framework can be induced, see e.g. Schulte im Walde (2006) for a discussion of the relationship between contextual similarity and theoretically defined verb classes and Čulo et al. (2008) for a comparison and discussion of semantic feature spaces. Another relevant distinction regarding DSM models concerns the way in which they are constructed. In what follows, I adopt the terminology of Baroni et al. (2014b) and refer to classical DSMs built by accumulating co-occurrence information from structured or unstructured data as "count''-DSMs, and to DSMs extracted with neural network architectures as "predict''-DSMs (see e.g. Mikolov et al. (2013)). At the quantitative level, count DSMs are high-dimensional while predict DSMs are low-dimensional. From a qualitative point of view, the dimensions of count-DSMs correspond to actual words, while the dimensions produced by predict-DSMs can be thought of as soft clusters of context items (Levy and Goldberg, 2014) that do not correspond to actual words.
The question for the conceptual underpinnings of verb meaning Whether or not the dimensions of a DSM model correspond to an actual word and thus are human-interpretable is irrelevant insofar as the adequacy of DSM representations is traditionally not determined by inspection of the DSM representation by itself but rather by evaluating the adequacy of a DSM representation against a gold standard (or a "Downstream Task") for a given clustering or classification problem. However, by focusing solely on the successful reproduction of a gold standard, Lenci (2014) concludes from a case study on structured DSM classification of Italian verbs, one may miss the right goal because one may well reproduce a given gold standard of classification while still there is "little understanding of the meaning components, i.e. the semantic features, relevant to analyze verb meaning". Importantly, as Lenci notes, the same difficulties with respect to the identification of the conceptual building blocks of word meaning arises for theoretical approaches to word meaning like the LCS framework. While in the LCS framework, too, "[t]he important theoretical construct is the notion of meaning component, not the notion of verb class" (Levin, 1993, p. 18), the identification of those conceptual elements involved in narrow-range lexical rules and the definition of semantically cohesive subclasses of verbs is the methodological blind spot of the LCS approach to verb meaning. For example, Van der Leek (1996) argues that the assumption that "[t]he subclasses of verbs that are eligible to enter into the conative alternation must signify a type of motion resulting in a type of contact." (Pinker, 2013, p. 123) is "purely stipulative" and that "there is no explanation why verbs that express motion and contact – and not even all of them – should enter into the alternation to the exclusion of verbs that do not" (Van der Leek, 1996, p. 365).

1.2 Preliminary work of the applicant

From this abridged presentation of the state of the art it appears that a theory of the conceptual underpinnings of word meaning, although of central importance to the deve-

lopment of a general theory of lexical semantics, remains an open research question in both theoretical and computational approaches. How are the recurrent conceptual building blocks of word meaning identified, represented and combined? This research question has led my research over the past few years. In this section, I summarize preliminary work in which I have addressed this question from the lexical-conceptual and distributional point of view.

Roots and concepts Given the pivotal role that conceptual information plays in the LCS theory of word meaning, the actual mechanisms that integrate conceptual meaning and formal templatic structure in the representation of a given word's meaning have not received the attention they deserve. One way towards integrating conceptual and formal aspects of word meaning that I have been exploring together with my colleagues in the project B4 of the SFB 732 is to adopt a more fine-grained view of the internal structure of words than the LCS theory assumes. More specifically, this more fine-grained view of word meaning rests on the assumption that words are formed from category-neutral, atomic and non-decomposable 'roots' which combine with features in the syntax to build larger linguistic elements (as in the morphological theory of Distributed Morphology (Halle and Marantz, 1993)). On the one hand, such a syntax-driven approach to word meaning makes it possible to render precise conceptual differences that are difficult to assess in LCS-style analyses, see e.g. Roßdeutscher and Kamp (2010) for a specification of the status of the direct object of non-core transitive verbs like malen ('to paint') which Levin (1999) stipulatively characterizes as 'pure constant arguments' (as opposed to 'structure arguments') or Pross (2015) for a syntax-based analysis of the conceptual meaning underlying so-called "Emission Verbs", a class of intransitive verbs which is notoriously difficult to explain on the basis of the unaccusativity hypothesis put forward in Perlmutter (1978). On the other, the syntax-based approach to word meaning requires us to handle both formal and conceptual aspects of meaning within the same component of the analysis, a point on which Pross and Roßdeutscher (2017) capitalize. In Pross and Roßdeutscher (2017), we investigate the relation between formal and conceptual aspects of word meaning, assuming that they are both syntactically represented. Because conceptual but not formal aspects of word meaning are sensitive to the type of direct objects, we argue with a case study on German denominal prefix and particle verbs like überdachen (to roof) that the proportion of formal and conceptual aspects of word meaning predicted by our syntactic analysis is reflected in the strength of the selection restrictions that these verbs impose on their direct objects. The predictions of our syntactic analysis are borne out empirically when selectional preference strength is modelled with the relative entropy of Germanet classes of direct objects as in Resnik (1996).

Dot-objects and concepts In Pross and Roßdeutscher (2017), we did not further qualify the source of the conceptual dimension of word meaning but rather used those 'primitive predicates' familiar from the LCS framework as 'placeholders' for the conceptual meanings

involved (but we noted that distributional semantics might be considered a way to characterize these placeholders). In doing so, we simply put aside the requirement imposed by the syntactic approach to word formation that both formal and conceptual aspects of word meaning must result from the compositional interpretation of the syntactic structure of a complex word to which root meaning is pivotal. On the one hand, root meaning feeds conceptual content into the interpretation of the syntactic context into which the root is inserted. On the other, root meaning determines which syntactic contexts are licit for insertion. Both these aspects of root meaning have been addressed in one go by making the assumption that roots have a certain conceptual (but no syntactic) category. For example, Marantz (1997) assumes that roots are conceptually categorized according to the lexical-semantic verb classes in Levin and Rappaport Hovav (1995) but in the next breath notes that "[t]he exact (semantic) categories for roots that predicts their varying behavior in nominal and verbal environments is not important [...] (although identifying these categories is of course essential to syntactic theory). The important point is that there are such categories" (Marantz, 1997, p. 216). Accounting for the meaning of roots in terms of classes defined by lexical-semantic templates may be intuitively plausible and – given the immense amount of groundwork in the LCS framework – easily accessible. But if the templatic classes of LCS are assumed to correspond to the atomic conceptual units of meaning, this simply fails to acknowledge that much more fine-grained conceptual distinctions within templatic verb classes – i.e. Levin's semantically cohesive subclasses – are key to the LCS framework. In Pross (2018), I proposed that one way to approach a more fine-grained theory of root meaning is to model root meaning in a form similar to how conceptual meaning is dealt with in the theory of dot-types (Pustejovsky, 1995; Asher, 2011). The type composition logic (TCL) developed in Asher (2011) distinguishes between two types of meaning: external and internal content. External content corresponds to the traditional model-theoretic extension of a word that determines its meaning at a certain point of evaluation. Internal content corresponds to a semantic object that encodes the content that expressions have by the way they are used and thus mirrors language users' systems of concepts. For example, the external meaning of the word book is a set of book entities at some world and time, while the internal meaning of book is given by the conceptual features we associate with book: a book is a physical object and an informational object. These conceptual features are represented as a structured type of concepts, a so-called dot-type as in (2).

(2) book $\rightarrow$ informational-object $\bullet$ physical-object

Each (sub)type of the conceptual aspects of a dot-object can serve as an extension of a word with which the dot-object is associated when that word is selected as the argument of a predicate which imposes restrictions on the conceptual type of its argument. That is, assuming that the verb to read selects for arguments associated with an informational

concept and to eat selects for arguments that are associated with an 'edible' concept, one can explain why (3-a) but not (3-b) is grammatical.

(3) a. Peter read the book.
b. *Peter ate the book.

Importantly, to deal with systematic semantic ambiguities (such as the related meaning of the verb to dance and the noun a dance) and accidental ambiguities (such as the unrelated meanings that the noun bank has), Asher (2011) assumes that dot-types are associated with syntactically uncategorized word stems rather than syntactically categorized words. Accordingly, because dot-types are not marked for syntactic category, they can be understood as representations of conceptual meaning, whereas the external meanings of dot-types in TCL serve as the meanings of words. While TCL provides a worked-out and systematic theory of how conceptual meaning relates to noun meaning, it is subject to the same methodological weaknesses as the LCS approach to verb meaning. In the LCS approach, the conceptual underpinnings of verb meaning are inferred from the observation of restrictions on argument structure alternations. In the theory of dot-objects, the conceptual underpinnings of noun meaning are inferred from the observation of restrictions on the acceptability of a given noun as the argument of a given verb. Both methods share the problem that they have to hypothesize conceptual restrictions rather than observing them directly from empirical data.

Distributions and concepts Pross et al. (2017) address the question for the conceptual building blocks of word meaning by using an unstructured predict-DSM approach to word meaning not only as a tool to reproduce an already established (human-crafted) gold standard but as way to explore previously unknown conceptual aspects of word meaning and thus as a genuine technique of lexical semantics on par with alternation-based approaches like the LCS framework. We show that when predict-DSM representations are rendered human-interpretable by approximation of the representation with its nearest neighbour words in the semantic vector space, the nearest neighbour characterization reflects conceptual commonalities between verbs similar to the narrow-range lexical rules of Pinker or Levin's semantically cohesive subclasses. Notably, these results tie in nicely with recent research in which DSM representations are considered as (albeit "very crude" ones (McNally and Boleda, 2017, p. 260)) representations for concepts, see e.g. Lenci (2008) for an overview.

Because the inspection of nearest neighbour characterizations of DSM representations with respect to their linguistically relevant internal conceptual structure is theoretically and methodologically basically terra incognita, Asher et al. (2016) being the only exception of which I am aware, I believe it is useful to illustrate the strategy of investigating conceptual structure with DSM representations pursued in Pross et al. (2017) with two examples. In Pross et al. (2017) we applied hierarchical clustering to predict-DSM rep-

resentations of über-prefixed verbs and compared the hierarchy output by the clustering with a classification of the same set of verbs into semantically cohesive verb classes, using observations at the syntax-semantics interface like argument structure alternations and case assignment as classification features. Manual inspection of the hierarchy output by the clustering algorithm showed that our lexical-conceptual classification was reproduced fairly well in that the verbs we assigned to the same lexical-conceptual class are by and large grouped together hierarchically. Moreover, when the uninterpretabale dimensions of the predict-DSM representations of the über-prefixed verbs are approximated by their ten nearest neighbours, the lexical entailments of the nearest neighbours provide a rough and schematic approximation of the conceptual underpinnings expected from a theoretical point of view. Consider e.g. the nearest neighbours of the "application" verb überkleben in (4).

(4) Ten nearest neighbours for "überkleben" (to paste over)
Aufkleber.N (sticker) bekleben.V (to glue on) Plakat.N (poster) Schriftzug.N (logo) Aufschrift.N (label) kleben.V (to glue) aufkleben.V (to affix) bedrucken.V (to print on) Aufdruck.N (imprint) prangen.V (to display)

The nearest neighbours in (4) provide a conceptually coherent topical characterization of the verb überkleben in that e.g. prototypical applicanda figure prominently, as well as other application verbs like bedrucken (to print on). This finding is pretty much in line with what standard interpretations of DSMs like Baroni et al. (2014a) contend. But the clustering experiment allowed for another, from a theoretical point of view more spectacular insight in that it produced the additional cluster in (5), where verbs which we classified differently in our lexical-conceptual approach are clustered together.

(5) "overpower''-cluster
überrollen (to roll over) überrennen (to overrun) überschwemmen (to overflow) überfluten (to flood) überfallen (to raid on sb.) überwältigen (to overpower) überkommen (to come over) übermüden (to overfatigue) überfahren (to overrun) überfressen (to overeat)
überschütten (to overwhelm sb. with sth.) überhäufen (to overheap)

When inspecting the cluster in (5), the question we asked ourselves is whether or not the cluster is conceptually coherent. Manual inspection of the nearest neighbours showed indeed that the verbs in (5) were not clustered together by accident but rather because they share a common conceptual core. As a prototypical example, consider the nearest neighbour characterization of the verb überrennen ('to overrun') in (6).

(6) Ten Nearest Neighbours for "überrennen" (to overrun):
Horde.N (mob) belagern.V (to besiege) Truppe.N (troop) Übermacht.N (supe-

riority) Streitmacht.N (army) einmarschieren.v (to invade) stürmen.V (to storm) erobern.V (to conquer) besiegen.V (to defeat) umzingeln.V (to surround)

What connects the nearest neighbours in (6) (and this observation generalizes to the other verbs in (5)) is that they share the lexical entailment of being related to unforeseeable overpowering instances of (natural) force exertion. Thus, it appears that DSM representations reflect conceptual commonalities between verbs similar to Levin's semantically cohesive subclasses, although nothing in the lexical-conceptual semantics of rennen or über indicates the possibility of a meaning shift like the one exemplified by überrennen. Consequently, Pross et al. (2017) conclude that DSM representations can not only be used to confirm those expectations about conceptual structure that emerge from a theoretical point of view but can also help in detecting conceptual aspects of verb meaning that are difficult if not impossible to target in frameworks of lexical semantics like LCS.

1.3 Project-related publications

1.3.1 Articles published by outlets with scientific quality assurance, book publications, and works accepted for publication but not yet published.

Tillmann Pross. What about lexical semantics if syntax is the only generative component of the grammar? A case study on word meaning in German. Natural Language and Linguistic Theory, 2017. Accepted with minor revisions.

1.3.2 Other publications

Tillmann Pross, Antje Roßdeutscher, Sebastian Padó, Gabriella Lapesa, and Max Kisselew. Integrating lexical-conceptual and distributional semantics: a case report. In Cremers, A. and van Gessel, T. and Roelofsen, F. (ed.): Proceedings of the 21st Amsterdam Colloquium, pp. 75 – 85, 2017.

1.4 Patents

1.4.1 Pending

Not applicable

1.4.2 Issued

Not applicable

Objectives and work programme

1.5 Anticipated total duration of the project

The project is planned for a duration of 18 Months. A funding by the DFG is requested for the whole duration of the project.

1.6 Objectives

General research hypothesis When putting together the pieces of the preliminary work of the applicant, and taking into account the open research question for the conceptual underpinnings of verb meaning, the general research hypothesis in (7) emerges.

(7) DSMs can be understood as representations of conceptual meaning. Thus, DSMs pave the way towards an empirically grounded theory of root meanings.

The big picture that (7) characterizes, and to which the proposed project aims to contribute, is schematized in (8).

(8) DSM →?→ Dot-Type/Conceptual Structure → Root Meaning → Word Meaning

The critical point of a theory of how conceptual meaning enters linguistic structures as characterized by (8), however, is the transition from DSM representations to conceptual structures in the spirit of the dot-objects advanced in TCL indicated by the question mark in (8). It is this transition from distributions to roots, with dot-objects serving as the mediating representation formalism with which the proposed project is concerned on an overarching level.

Objective of the project It is often noted that one of the main advantages of DSM representations over approaches to conceptual meaning like LCS or TCL is that DSM representations can be automatically induced, are easy to construct, empirically well-founded and bear psychological plausibility. Thus, it is not far to seek a combination of the characterization of conceptual content provided by DSM representations with a symbolic representation of conceptual structures like the dot-objects of TCL. In fact, Asher et al. (2016) present a case study on noun-adjective combinations where dot-objects are related to DSM representations by interpreting the ten nearest neighbours of a DSM representation of a word as the subtypes of the dot-object associated with that word. However, although Asher et al. (2016) show that the nearest neighbour characerization is in principle able to capture important semantic properties of adjective-noun combinations like subsectivity and intersectivity or meaning shifts (when the nearest neighbours of an adjective in isolation are compared to the nearest neighbours of an adjective in a nominal context), they leave open the most important step of the transition from nearest neighbour approximations of DSM representations to conceptual structures. In order to systematically (let alone

automatically) translate DSM representations into conceptual structures of the type represented by dot-objects, "we need a separate process that clusters the predicates into different coherent internal meanings" (Asher et al., 2016, p.718), which "as far as we know, is not yet feasible". It is exactly this gap in the transition from (continuous) DSM representations into the purely symbolic environment of root-based lexical (and subsequently formal) semantics that the proposed project aims to address, building on the methods used and insights made in previous work of the applicant. The other novel contribution of the proposed project is that whereas previous work like Asher et al. (2016) or McNally and Boleda (2017) considers only the conceptual dimension of adjective-noun combinations, the proposed project examines the conceptual underpinnings of verb meaning.

The objective of the project is to research whether and how symbolic representations of conceptual structures of the type proposed in TCL can be derived from DSM representations as part of the general view of how conceptual meaning enters linguistic predication that is depicted in (8). Thus, the project aims at investigating the "translation" of DSM representations into dot-objects by rendering predict-DSM representations transparent for human interpretation and inspecting these transparent human interpretations for conceptually coherent subcomponents. In turn, if dot-objects are understood as root meanings and root meanings are the atomic conceptual building blocks of meaningful linguistic structures, this project aims to examine how robust linguistic expectations about conceptual structures are when confronted with distributional characterizations of conceptual structures. Figuratively speaking, the project investigates whether there is a systematic, theoretically informed and computationally supported way to perform the drastic "dimension reduction" that is required to convert a 300-dimensional predict-DSM representation of a word into a dot-object with, say, 4 subtypes such that the dot-object can be understood as a theoretically reasonable and sensible representation of the root meaning of that word. The project thus intends to make some steps towards bridging an important gap between the state of the art in theoretical and computational linguistics. On an overarching level, it aims at showing that a combination of theoretical, lexical-conceptual and computational, usage-based approaches to verb meaning may pave the way towards an empirically grounded and theoretically sound theory of verb meaning in its entirety.

Scope of Investigation As the project is devoted to conceptual groundwork, its scope and aims are naturally limited. First, the project sticks by and large to the investigation of the conceptual meaning of morphologically simple verbs in English and German. It does not consider those questions that arise when meanings are combined in phrases and sentences, i.e. the question for the distinction between "referential" and "conceptual" affordances, in the terminology of McNally and Boleda (2017), or "rigid" and "holistic" meaning composition in the terminology of Pross et al. (2017) or "external" and "internal" meaning in Asher et al. (2016). Second, I consider a selection of data that is small enough to allow for a comprehensive qualitative investigation and for which relatively precise

theoretical predictions have been made concerning their conceptual underpinnings. Third, the project sticks by and large to already established methodologies from computational linguistics. That is, I do not aim at advancing new or better suited algorithms for the extraction of DSM representations from corpora but will draw on established out-of-the-box algorithms and use pre-trained models (if available). Fourth, I aim at a qualitative characterization of the relation between DSM representations, dot-objects and roots as the basis for a future quantitative assessment. I believe that developing algorithms that map DSM representations to structured concepts which in turn encode syntactically relevant information requires at first a profound investigation of what the structured concepts are like that DSM representations may or may not encode. Keeping these limitations in mind, the proposed project will be pursued within the three work packages (WPs) detailed in the next section.

2 Work programme incl. proposed research methods

2.1 WP1: Quantitative Characterization, Data Generation (2 Months)

DSM representations typically have several hundred or several thousand dimensions (depending on whether a count or predict DSM is used). Dot-types are built from only a few (say, less than five) primitive concepts. Thus, the research problem to which WP1 is devoted is to reduce the dimensionality of DSM representations quantitatively. Statistical approaches to natural language meaning are highly empirical and often involve several iterations between modelling, experimentation and interpretation. Pross et al. (2017) is the result of several such iterations, where the final computational setting of the investigation is the one that turned out to deliver the best results with respect to the given task. We found that the inherent dimensionality reduction performed by continuous bag of words (CBOW) predict-DSMs works best with respect to the manual identification of salient meaning components when the dimensions of the DSM representation are approximated with nearest neighbours (e.g. as compared to a combination of count-DSMs and dimensions reduction through singular value decomposition). We computed the nearest neighbours for each of the extracted vectors V by using the dot-product of V and all other vectors in the vector space (as in Levy and Goldberg (2014)) as a measure of proximity, because this method turned out to be most successful with respect to the interpretability of the approximated continuous representations (e.g. as compared to cosine similarity, the proximity measure employed in Asher et al. (2016)). For the clustering step of the case study in Pross et al. (2017), hierarchical agglomerative clustering with average linkage turned out to deliver the best results with respect to the production of conceptually coherent clusters (e.g. as compared to k-means clustering).

A natural question that arises from these specific parameter settings is of course whether

the observations made when manually inspecting the nearest neighbour characterizations are stable across different extraction methods for DSMs, dimensionality reduction methods, similarity measures and clustering algorithms. For the proposed project, I take this question into account by generating interpretable DSM representations for verbs with a range of methods. I plan to use the two algorithms for the extraction of the predict-DSM (CBOW vs. skip-gram) proposed in Mikolov et al. (2013) with different (hyper-)parameter settings like window sizes. As in Pross et al. (2017), I plan to compute the nearest neighbours of the dense DSM representations using the dot-product method. Besides the nearest neighbours approach to rendering predict-DSM representations interpretable, I also plan to derive interpretable predict-DSM characterizations using non-negative sparse embeddings of predict-DSMs as described in Faruqui et al. (2015). To round out the picture, I will examine a more experimental way of rendering DSM representations interpretable for which we carried out some initial experiments in preparatory work for Pross et al. (2017). The idea is to encode the representations of words in the vector space with an n-hot representation of the dimensions with the highest and lowest values. The similarity of a given word can then be approximated by calculating the similarity of the n-hot encoding of that word with the n-hot encodings of the dimensions of all other words in the vector space. Finally, as a rough computational approximation of the idea that DSM representations have an internal structure of separable conceptually coherent meaning components comparable to the aspects of a dot-object, I will cluster the predict-DSM representations of the elements of the interpretable representations by using a range of different methods, e.g. k-means and hierarchical clustering with variable parameter settings.
Milestone: The output of WP1 is (a) a set of interpretable DSM representations derived by different methods for the data set detailed in WP2 and (b) for each of the interpretable DSM representations, clusterings of the elements of this representation of the data set detailed in WP2.

2.2 WP2: Qualitative Investigation (12 Months)

In WP2, I examine the main research hypothesis of the proposed project with a case study on intransitive verbs. From a theoretical point of view, intransitive verbs are interesting because according to the so-called unaccusative hypothesis put forward in Perlmutter (1978) there are two types of intransitive verbs. If the grammatical structure of a transitive verb relates a grammatical subject to a grammatical object, then the grammatical structure of unergative verbs like to laugh has a grammatical subject but no grammatical object and the grammatical structure of unaccusative verbs like to stumble has a grammatical object but no grammatical subject. In German the distinction between unergative and unaccusative verbs is syntactically represented. For example, German unergative verbs like lachen (to laugh) appear in impersonal passives while unaccusative verbs like sterben

(to die) do not. Or, unergative verbs like lachen select the perfect auxiliary haben (have) while unaccusative verbs like sterben select sein (be).

What makes intransitive verbs particularly interesting as a subject of study with respect to the goals of the proposed project is that the grammatical distinction between unergative and unaccusative verbs correlates with a distinction in the understood conceptual underpinnings: "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." (Dowty, 1991, p. 605). While the intuition that unergative and unaccusative verbs encode a fundamental conceptual distinction between say, Agent and Patient, the exact definition of the concepts relevant to the unaccusative/unergative distinction is an unsettled issue, see e.g. Pross (2015) for a comparison of the incompatible views of Levin and Rappaport Hovav (1995) and Reinhart (2002) (and Rappaport Hovav and Levin (2000) for a general overview). The fundamental role that the unaccusativity hypothesis plays in modern theoretical linguistics (see e.g. Alexiadou et al. (2004)) and the simplicity of the basic intuition concerning the conceptual difference between unaccusative and unergative verbs as well as the lack of a clear consensus about how this intuition should be rendered precise makes intransitive verbs an ideal testing ground for the general research hypothesis pursued in the proposed project. This is because the hypothesis that the conceptual structure of a verb's meaning is reflected in the DSM representations of that verb provides a novel perspective on the much debated question for the conceptual underpinnings of the unaccusativity hypothesis that can be summarized with the two main general research questions of WP2 in (9).

(9) a. How robust are the intuitions standardly associated with the conceptual difference between unaccusative and unergative verbs when confronted with the theoretically unbiased, empirically grounded and psychologically plausible conceptual characterization of these verbs provided by interpretable DSM representations?

b. Can DSM representations provide novel insights into the conceptual underpinnings of intransitives (if DSM representations are not only used as a tool for replication but also as an explorative method, as reported in Pross et al. (2017))?

I assess these questions by examining 10 items from English and German for each of the word classes in (10).

(10) a. morphologically simple unergative verbs (run, work / laufen, arbeiten)

b. morphologically simple unaccusative verbs (stumble, die / stolpern, sterben)

c. unergative incremental theme verbs that enter the unspecified object alternation (eat, paint / essen, malen)

d. unaccusative anticausative verbs (break, melt / zerbrechen, schmelzen)

The first goal of WP2 is to identify clues, pointers and indicators for the encoding of conceptual structures in interpretable DSM representations and to formulate general patterns, templates and schemas in the form of rules of thumb – a "proto-theory" for the conceptual interpretation of DSM representations. The second goal of WP2 is to systematically confront the hypothesis that unaccusative and unergative verbs are conceptually distinct with the characterizations of these verbs in terms of interpretable DSM representations. On the basis of these research goals, WP2 aims at a characterization of the conceptual underpinnings of intransitives by examining in more detail how DSM representations encode conceptual structures along the research questions in (11).

(11) a. Can the nearest neighbours/sparsified representations of the verbs in (10) be grouped manually into semantically cohesive clusters? Do these clusters correspond to the output of the automatic clustering?
b. If semantically cohesive clusters can be observed, what kind of concepts do they possibly represent? Are the concepts related to those that are standardly assumed in the theoretical literature on unaccusativity?
c. If there are semantically cohesive clusters, are these clusters stable across different extraction/clustering algorithms and (hyper)parameter settings?

In the final part of WP2, I compare the DSM-derived conceptual characterizations for the English and German verbs in (10), to asses the degree to which DSM-derived representations are language-specific. A comparison of English and German is also interesting from a theoretical point of view, as the unaccusativity hypothesis is often understood as a cross-linguistic generalization about the deep structure of grammatical relations (see e.g. Levin and Rappaport Hovav (1995)). Nevertheless, languages differ with respect to the degree the unergative/unaccusative distinction correlates with overt morpho-syntactic markers. E.g., whereas in German indicators like auxiliary selection, impersonal passives or prenominal participles (see e.g. Grewendorf (1989) for discussion) can be used to distinguish unergative from unaccusative verbs, unaccusativity in English has been argued to mainly correlate with semantic properties like e.g. the licensing of resultative constructions, see Levin and Rappaport Hovav (1995). I will address these more general issues concerning the language-independency of DSM-derived conceptual structures with the research question in (12).

(12) Are there language-specific differences between the DSM-derived conceptual characterization of intransitives in German and English? If there are such differences, how do these differences relate to an understanding of the unaccusativity hypothesis as a generalization about the deep grammatical structure of verbs?

Milestone: The envisaged output of WP2 is a lexicon-like list of DSM-derived conceptual structures for the verbs in (10), using the semantic objects of dot-types in TCL as a representation formalism.

2.3 WP3: Quantitative evaluation (4 Months)

To quantify the extent to which the conceptual clusters identified in WP2 are conceptually coherent, I propose to evaluate the findings of WP2 with a word intrusion experiment (Chang et al., 2009). In a word intrusion experiment, humans are asked to single out one word from a set of words on the basis of conceptual incoherence. For example, participants would be asked to single out the intruder word from the automatically generated nearest neighbour characterization of überrennen in (13). If a statistically significant number of participants selects Pizza as the intruder word in (13), this allows to quantitatively determine the characterization of überrennen as being conceptually coherent and semantically cohesive.

(13) Nine Nearest Neighbours for "überrennen" (to overrun) and One Intruder
Horde.N (mob) belagern.V (to besiege) Truppe.N (troop) Pizza.N (pizza) Streitmacht.N (army) einmarschieren.v (to invade) stürmen.V (to storm) erobern.V (to conquer) besiegen.V (to defeat) umzingeln.V (to surround)

The main challenge when carrying out a word intrusion experiment concerns the selection of plausible intruders. This is because the experiment designer must identify the middle ground between selecting intruders that are totally conceptually incoherent with the cluster to be assessed for coherence (like Pizza in (13)) and selecting intruders that are totally conceptually coherent with the cluster to be assessed (e.g., if the word erobern were the intruder in (13)). In both these extreme cases, the experiment becomes trivial and the same is true for a simple random choice of the intruder. In the literature (e.g. Murphy et al. (2012); Faruqui et al. (2015)), the problem of intruder choice is approached by selecting intruders from a set of words that (a) have a low probability to occur in the cluster to be assessed for conceptual coherence but (b) have a high probability to occur in some other conceptual cluster (that is assessed in the experiment). Besides this established approach to intruder selection, I also plan to test a more experimental "gradual" approach to intruder selection. The idea is to incrementally increase the conceptual coherence of the intruder with the cluster of words to be assessed until the point when the intruder cannot be identified any longer by the participants of the experiment, and, if necessary, compare the points at which intruders become coherent across conditions.
Milestone: The envisaged output of WP3 is a quantitative evaluation of the findings of WP3 using a word intrusion experiment.

3 Bibliography

Artemis Alexiadou, Elena Anagnostopoulou, and Martin Everaert, editors. The Unaccusativity Puzzle: Explorations of the Syntax-Lexicon Interface. Cambridge University Press, 2004.

Nicholas Asher. Lexical Meaning in Context: A Web of Words. Cambridge University Press, 2011.

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

Marco Baroni and Alessandro Lenci. Distributional memory: A general framework for corpus-based semantics. Computational Linguistics, 36(4):673—721, 2010.

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

Marco Baroni, Georgiana Dinu, and Germán Kruszewski. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pages 238 — 247, 2014b.

Manfred Bierwisch. Semantic form as interface. In Andreas Späth, editor, Interfaces and Interface Conditions, pages 1–32. de Gruyter, Berlin, 2007.

Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L. Boyd-Graber, and Blei. David M. Reading tea leaves: How humans interpret topic models. In Proceedings of NIPS, 2009.

Stephen Clark. Vector space models of lexical meaning. In Shalom Lappin and Chris Fox, editors, The Handbook of Contemporary Semantic Theory, pages 493 – 522. Wiley Blackwell, 2 edition, 2015.

Oliver Čulo, Katrin Erk, Sebastian Padó, and Sabine Schulte im Walde. Comparing and combining semantic verb classifications. Language Resources and Evaluation, 42(3): 265–291, Sep 2008. ISSN 1574-0218. doi: 10.1007/s10579-008-9070-z.

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

Manaal Faruqui, Yulia Tsvetkov, Dani Yogatama, Chris Dyer, and Noah A. Smith. Sparse overcomplete word vector representations. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pages 1491—1500, Beijing, China, 2015.

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

Mohamed Guerssel, Kenneth. Hale, Margaret Laughren, Beth Levin, and Josie White Eagle. A cross-linguistic study of transitivity alternations. In CLS 21: Papers from the Parasession on Causatives and Agentivity., volume 2, pages 48–63. Chicago Linguistic Society, 1985.

Morris Halle and Alec Marantz. Distributed morphology and the pieces of inflection. In Kenneth Hale and Samuel Jay Kaiser, editors, The View from Building 20. Essays in Linguistics in Honor of Sylvian Bromberger., pages 111 – 176. MIT Press, 1993.

Alessandro Lenci. Distributional semantics in linguistic and cognitive research. Italian Journal of Linguistics, 20(1):1 – 31, 2008.

Alessandro Lenci. Carving verb classes from corpora. In Raffaele Simone and Francesca Masini, editors, Word Classes: Nature, typology and representations, pages 17 – 36. John Benjamins, 2014.

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

Beth Levin. Objecthood. an event structure perspective. In Proceedings of CLS 35, pages 223–47. Chicago Linguistic Society, 1999.

Beth Levin and Steven Pinker. Introduction. In Lexica & Conceptual Semantics, pages 1 – 8. Blackwell, 1991.

Beth Levin and Malka Rappaport Hovav. Unaccusativity at the syntax-semantics interface. MIT Press, 1995.

Omer Levy and Yoav Goldberg. Neural word embedding as implicit matrix factorization. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 27, pages 2177–2185. Curran Associates, Inc., 2014.

Alec Marantz. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In University of Pennsylvania Working Papers in Linguistics, volume 4, Issue 2, Article 4, 1997.

Louise McNally and Gemma Boleda. Conceptual vs. Referential Affordance in Concept Composition, pages 245–267. Springer International Publishing, Cham, 2017.

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

Brian Murphy, Partha Pratim Talukdar, and Tom Mitchell. Learning effective and interpretable semantic models using non-negative sparse embedding. In International Conference on Computational Linguistics (COLING 2012), pages 1933–1949, 2012.

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

Steven Pinker. Learnability and Cognition: The Acquisition of Argument Structure. New edition. MIT Press, 2013.

Tillmann Pross. Mono-eventive verbs of emission and their bi-eventive nominalizations. In Thui Buy and Denis Özyıldız, editors, Proceedings of NELS 45, volume 1, pages 257–266, Amherst, MA, 2015. GLSA.

Tillmann Pross. What about lexical semantics if syntax is the only generative component of the grammar? a case study on word meaning in german. Natural Language and Linguistic Theory, 2018. doi: https://doi.org/10.1007/s11049-018-9410-7.

Tillmann Pross and Antje Roßdeutscher. Measuring out the relation between formal and conceptual semantics. In K. Balogh and W. Petersen, editors, Bridging formal and conceptual semantics. Selected papers of the workshop on bridging conceptual and formal semantics (BRIDGE-14), pages 119–149. Düsseldorf University Press, 2017.

Tillmann Pross, Antje Roßdeutscher, Sebastian Padó, Gabriella Lapesa, and Max Kisselew. Integrating lexical-conceptual and distributional semantics: a case report. In Proceedings of the Amsterdam Colloquium 2017, pages 75–85, 2017.

James Pustejovsky. The Generative Lexicon. MIT Press, 1995.

Malka Rappaport Hovav and Beth Levin. Classifying single argument verbs. In Peter Coopmans, Martin Everaert, and Jane Grimshaw, editors, Lexical Specification and Insertion, volume 197 of Current Issues in Linguistic Theory, pages 269 – 304. John Benjamins, 2000.

Tanya Reinhart. The theta system - an overview. Theoretical Linguistics, 28:229–290, 2002.

Philip Resnik. Selectional constraints: an information-theoretic model and its computational realization. Cognition, 61:127–159, 1996.

Antje Roßdeutscher and Hans Kamp. Syntactic and semantic constraints on the formation and interpretation of ung-Nouns. In Artemis Alexiadou and Monika Rathert, editors, Nominalisations across Languages and Frameworks. de Gruyter Mouton, Berlin, 2010.

Sabine Schulte im Walde. Experiments on the Automatic Induction of {G}erman Semantic Verb Classes. Computational Linguistics, 32(2):159–194, 2006.

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

Frederike Van der Leek. The english conative construction: A compositional account. In CLS 32: Papers from the Main Session, volume 32, pages 363–378. Chicago Linguistic Society, 1996.