# Integrating lexical-conceptual and distributional semantics: a case report. *

Tillmann Pross, Antje Roßdeutscher, Sebastian Padó, Gabriella Lapesa, and Max Kisselew

University of Stuttgart

**Abstract**

By means of a case study on German verbs prefixed with the preposition *über* ('over') we compare alternation-based lexical-conceptual and usage-based distributional approaches to verb meaning. Our investigation supports the view that when distributional vectors are rendered human-interpretable by approximation of their representation with its nearest neighbour words in the semantic vector space, they reflect conceptual commonalities between verbs similar to those targeted in lexical-conceptual semantics. Moreover, our case study shows that distributional representations reveal conceptual features of verb meaning that are difficult if not impossible to detect and represent in theoretical frameworks of lexical semantics and thus that a general theory of word meaning requires a combination and complementation of lexical and distributional methods.

## 1 Introduction

A general theory of lexical representation is key to a compositional theory of the meaning of supralexical linguistic expressions. On these premises, the present paper investigates the relation between two approaches to word meaning: alternation-based lexical-conceptual semantics and usage-based distributional 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" [7]. One way to represent these syntactically relevant components of meaning is to decompose a verb's meaning into a fixed set of primitive predicates and constants from a limited set of semantic types. 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', and in which a constant names the state (e.g. 'broken') filling the second argument of 'become'. 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*, it has been proposed that the relevant distinction is of a conceptual nature. In the terminology of [10], 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 syntax can be sensitive to. That is, syntactic evidence not only provides a characterization of the general "templatic" aspects of verb meaning but also of the narrow-range constraints on the usage of a particular verb. As [6] shows impressively, when we extend the search for such syntactically represented conceptual distinctions to a wider range of verbs and constructions,

*The research reported in this paper has been supported by a DFG grant to the projects B4 (Pross, Roßdeutscher) and B9 (Padó, Lapesa, Kisselew) of the Collaborative Research Centre 732 "Incremental Specification in Context."

a systematic and fine-grained lexical-conceptual classification of verb meaning can be induced. We refer to this particular alternation-based approach of verb meaning in the following as the lexical-conceptual structure (LCS) approach to verb meaning.

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" [12]. 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. [3]): *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 [1], where DSM representations are not harvested from 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. [11] for a discussion of the relationship between contextual similarity and theoretically defined verb classes. Another relevant distinction regarding DSM models concerns the way in which they are constructed. In what follows, we 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. 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 [8] that do not correspond to actual words. However, whether or not the dimensions of a DSM model correspond to an actual word is insofar irrelevant 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. But by focusing solely on the successful reproduction of a gold standard, [5] 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, 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, as the identification of those conceptual elements involved in narrow-range lexical rules and the definition of semantically cohesive subclasses of verbs are the methodological blind spot of the LCS approach to verb meaning. For example, [13] argues that the assumption that contact and motion are required for a verb to enter the conative construction are "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".

We address the question for the conceptual building blocks of word meaning by using the 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 resulting characterization reflects conceptual commonalities between verbs similar to the narrow-range lexical rules of Pinker or Levin's semantically cohesive subclasses and in fact reveals conceptual features of verb meaning that are difficult if not impossible to detect and represent in frameworks of lexical semantics like LCS. We develop our argument by comparing a classification of 80 German verbs prefixed with the preposition *über* ('over') into semantically cohesive verb classes à la Levin with the output of an unsupervised clustering of the same set of *über*-verbs (section 2). Second, we argue that rendering DSM representations transparent is not only highly diagnostic for word meaning but even more so for more complex cases of meaning composition (section 3). Adopting an additive model of the composition of DSM representations, we show that rendering transparent the difference vector that results from subtracting the DSM representation of a base verb from the DSM representation of an *über* prefixed-verb reveals insights into the conceptual underpinnings and effects of the process of prefixation like meaning shifts which, although linguistically reflected, standardly escape the attention of lexical semanticists. Section 4 concludes.

## 2 Simple meaning spaces

The basic use of *über* ('over') is as a preposition with two distinct meanings. Depending on the aspectual class of the matrix verb, an *über*-PP can refer to the direction of the motion of an accusative reference object as in (1) or to the location of a dative reference object as in (2).

(1) Der Mann sprang über den Zaun.
the man jump over the.ACC fence
"The man jumped over the fence"

(2) Das Bild hing über der Tür.
the painting hang over the.DAT door
'The painting hung above the door'

German has a productive mechanism of word formation by affixation of prepositional elements like *über* to a base verb. In the following, we distinguish four lexical-conceptual classes of German *über*-affixed verbs by considering the participation of these verbs in locative alternations, the licensing of PP complements and case assignment. First, when *über* is affixed to a verb as in (3), the derived verb describes a movement ACROSS some obstacle. As (1) shows, a PP complement construction with *über* is licensed with motion verbs like *springen*.

(3) Der Mann übersprang den Zaun.
the man over-PRFX.jump the.ACC fence
'The man jumped over the fence'

Second, when *über* is affixed to change of possession verbs like *geben* ('to give'), the prefixed verb describes the *transfer* of an object x from A to B as in (4). The argument marked with dative case identifies the location at which the transferred object x ends up. No *über* PP-complement construction is possible with the base verb (5).

(4) Er übergab ihr den Brief
he over-PRFX.give her.DAT the.ACC letter
'He handed her over the letter'

(5) *Er gab den Brief über sie.
he give the letter over her

A third class of *über*-affixed verbs describes the APPLICATION of an object to another object as in (6-a). This class of APPLICATION verbs is distinguished from the ACROSS class by participation in a locative alternation as in (6-a)/(6-b).

(6) a. Peter überklebte den Kratzer mit einem Aufkleber.
Peter over-PRFX.glue the scratch with a sticker.
'Peter over-pasted the scratch with a sticker'

b. Peter klebte den Aufkleber über den Kratzer.
Peter paste the sticker over.PREP the scratch.
'Peter pasted the sticker over the scratch'

Fourth, *über*-affixation of a verb can also be used to describe that the event denoted by the verb exceeds a certain contextual standard on a SCALE provided by the base verb, see (7). No PP-complementation with *über* is possible for the SCALE-class and the direct object receives accusative case (8).

(7) Er überbewertete die Aktie.
he over-PRFX.value the.ACC share
'He overvalued the share'

(8) *Er bewertete über die Aktie
he value over the share

We assigned up to 20 *über*-prefixed verbs to each of the four lexical-conceptual classes identified in the previous section and extracted distributional vectors with 300 dimensions for the *über*-prefixed verbs and their morphologically and semantically related base verbs using the CBOW model proposed by [9] with a symmetric 5-word window. The vectors were extracted from SdeWac [4], a web corpus created from a subset of the DeWaC corpus. It contains about 45m sentences selected to be well-formed sentences. We use an unstructured DSM because these models are the simplest possible ones, make the fewest assumptions, and we were interested in assessing the topic-oriented perspective that they provide (rather than the relationally-oriented perspective of structured DSMs). We then computed pair-wise cosine similarity between the distributional vectors. We then tried to establish a hierarchy among the computed pairwise similarities with the hierarchical agglomerative clustering algorithm from the SciPy package using the unweighted pair group method with arithmetic mean as linkage algorithm. Manual inspection of the hierarchy output by the clustering showed that our lexical-conceptual classification is reproduced fairly well in that the verbs from the TRANSFER class (t) in (9), the SCALE class (s) in (10), the ACROSS class (a) in (11) and the APPLICATION class (ap) in (12) are by and large grouped together hierarchically. Certainly, each of the clusters contains some outliers, but closer inspections shows that these outliers are mainly due to errors in the preprocessing or ambiguities. This is a remarkable result, insofar as the underlying DSM is unstructured, whereas in computational linguistics verb classes are standardly reproduced with structured DSMs.

(9) TRANSFER übergehen (pass s.th.over) (a); übereignen (convey) (t); überführen (lead across) (a); übernehmen (take over) (t); überlassen (let s.o. s.th. for use) (t); überantworten (pass repsonsibility) (t); übersenden (send) (t); übermitteln (transfer)(t); überreichen (hand over) (t); übergeben (hand over) (t); überweisen (trans-scribe) (t);

(10) SCALE überstimmen (outvote) (s); überrepräsentieren (overrepresent) (s); überspielen (copy) (t); überhören (miss s.th.) (a); überreizen (overexite) (s); überfordern (overstrain) (s); überstrapazieren (overstrain) (s); übertreiben (overdo) (s) ; übersteigern (surmount) (s); überzeichnen (make burlesque) (s); überdrehen (overwind) (s); überspitzen (exaggerate) (s); überhöhen (inflate) (s); überladen (overload) (s); überfrachten (overcharge) (s); überschätzen (overestimate) (s); überbewerten (overrate)(s); übersehen (overlook) (a); überwiegen (outweigh) (s); überbuchen (overbook) (s);

(11) across übersetzen (translate) (t); überliefern (pass down) (t); überschreiben (transfer) (t); überlesen (skip) (a); überblättern (page over) (a); überfliegen (fly across) (a); überarbeiten (overwork) (s); überschreiten (overstep) (a); übertreten (cross) (a); überspringen (jump over) (a); überschauen (survey, overlook) (a); überkreuzen (cross) (a);

(12) application überhängen (cover by hanging s.th.) (ap); überstreuen (cover with sprinkles) (ap); überstäuben (cover with dust) (ap); übergießen (douse) (ap); übersprühen (cover by spraying) (ap); überstreichen (cover with paint) (ap); übermalen (cover by painting) (ap); überkleben (paste over)(ap); überziehen (cover with a coat) (s); übertünchen (cover with whitewash) (ap); überdecken (cover) (ap); überlagern (overlay, interfere) (ap); überbauen (build s.th. across s.th.) (a); überklettern (climb over) (a); überwachsen (overgrow) (ap); übersäen (reseed) (ap); überragen (tower above)(a);

But the clustering allowed for an even more interesting insight, as it gave rise to the additional fifth cluster in (13), where verbs which we classified differently in our lexical-conceptual approach are clustered together.

(13) overpower überrollen (overrun) (a); überrennen (overrun)(a); überschwemmen (flood, drown) (ap); überfluten (deluge) (ap); überfallen (attack) (s); überwältigen (overwhelm) (s); überkommen (be assailed by sth.) (trans); übermüden (overfatique) (s); überfahren (knock down) (a); überfressen (overeat) (s); überschütten (spill s.th. on s.o.) (ap); überhäufen (heap on) (ap);

If, as is customary in computational linguistics, the quality of the clustering would be measured in terms of predicting the gold standard provided by our four hand-crafted lexical-conceptual classes, then we would have to conclude from (13) that the parameter settings of our clustering algorithm should be revised to achieve a higher precision. But closer inspection of the verbs in the fifth cluster suggests that there may be another option to interpret the clustering result: Maybe the additional cluster did not come about by accident but identifies an additional class of *über*-verbs which we were not able to detect with the admittedly simplistic lexical-conceptual diagnostic tools we employed. Because predict-DSM representations cannot be assessed to find out whether the fifth cluster came about by accident (and thus the algorithm is wrong) or is semantically cohesive (and thus the gold standard is wrong) we approximated the vector representations of the *über*-verbs in the fifth cluster with their "nearest neighbours" (where proximity in space of two vectors is identified by their dot product as in [8]) to determine the ten words nearest in the semantic vector space to the target word. Consider the base verb *rennen* ('to run') (14) and the derived verb *überrennen* ('to overrun') (15).

(14) *rennen* (to run) base

| | | | | | | |
|---|---|---|---|---|---|---|
| springen.V | schnappen.V | zurennen.V | hüpfen.V | wegrennen.V | schreien.V | brüllen.V |
| jump | snap | towards-run | hop | run-away | scream | yell |
| schleichen.V | aufspringen.V | schreiend.A | | | | |
| creep | jump-up | screaming | | | | |

(15) *überrennen* (to overrun) derived

| | | | | | |
|---|---|---|---|---|---|
| Horde.N | belagern.V | Truppe.N | Übermacht.N | Streitmacht.N | einmarschieren.v |
| hord | besiege | troop | superiority | force | invade |
| stürmen.V | erobern.V | besiegen.V | umzingeln.V | | |
| assault | conquer | defeat | surround | | |

What the representation for *(über)rennen* shows, and this generalizes to the verbs that were clustered together in (13), is that these verbs were not clustered together by accident but rather because they share a common conceptual core. The *über*-prefixed verbs describe unforeseeable events of overpowering instances of (natural) forces exertion. Interestingly, nothing in the lexical semantics of *rennen* or *über* (at least according to the standards of lexical-conceptual semantics) indicates the possibility of such a meaning shift through *über*-prefixation. Although apparently trivial, the observation that the nearest neighbour characterizations which can render opaque DSM representations interpretable by humans encode a certain kind of lexical-conceptual knowledge in the sense of Levin and Pinker has not been made in the literature before. One reason for this may be, as already mentioned, that DSM representations are standardly evaluated with respect to a gold standard. Gold standards are tied to specific purposes and hypotheses, whereas what we aim at doing is exploratory work, i.e. to try to give a linguistic interpretation to the information encoded in a DSM. Moreover, making DSMs transparent indicates an advantage of using an unstructured DSM, because the nearest neighbours of a given vector are topical in nature and do not require similarity with regard to the fillers of specific syntactic positions (e.g. direct objects). In this manner, they capture more abstract and general conceptual features of the semantic space, as indicated e.g. by the verbs *belagern* ('to besiege') and *umzingeln* ('to surround') in (15).

## 3 Complex meaning spaces

We suggested in the previous section that DSM representations encode aspects of word meaning that are difficult to target by means of grammaticality judgements at the syntax-semantics interface as in the LCS-framework. What kind of observations are we to expect for the composition of DSM representations? To approach this question, we adopted an additive model of the composition of DSM representations [2], and represented the meaning shift that results from the composition of a base verb with its prefix by the difference between the base verb vector and the prefix verb vector. Using the same method of nearest neighbour approximation as in the previous section, we rendered transparent the "shift" vector that results from subtracting the DSM representation of a base verb from the DSM representation of the corresponding *über*-prefixed verb. Thus, we did not try to learn one general DSM representation of the prefix *über* (because a general DSM representation will smooth out the meaning of *über*) but calculated for each pair of observed base and derived verb the specific "'surplus" that *über* makes to the construction. We then investigated the question whether a general semantic function of *über*-prefixation can be induced from the idiosyncractic meaning that our additive model of DSM representations assigns to *über* in a specific construction. Consider first (16).

(16) *kleben* (to glue) BASE

| | | | | |
|---|---|---|---|---|
| **aufkleben.V** | ausschneiden.V | Klebeband.N | festkleben.V | **bekleben.V** |
| glue.on.PRTC.glue | out.PRTC.cut | tape | fix.glue | be.PRXF.glue |
| verkleben.V | tropfen.V | ankleben.V | bemalen.V | abwischen.V |
| fix | drop | on.glue | be.PRFX.paint | wipe-off |

(17) *überkleben* (to cover) DERIVED

| | | | | | |
|---|---|---|---|---|---|
| Aufkleber.N | **bekleben.V** | Plakat.N | Schriftzug.N | Aufschrift.N | **kleben.V** |
| sticker | be.PRXF.glue | poster | letters | label | glue |
| **aufkleben.V** | bedrucken.V | Aufdruck.N | prangen.V | | |
| on.PRTC.glue | be-print | logo | be-respleshdent | | |

(18) *über* (over) SHIFT

| | | | | | | |
|---|---|---|---|---|---|---|
| vorgenommen.A | Bundesarchiv.N | Bürgerbegehren.N | Rüstungsexport.N | | | |
| planned | federal-archive | petition-referendum | export-of-arms | | | |
| Freiheitsstrafe.N | Umbenennung.N | erfolgt.A | Kürzung.N | staatlich.A | irreführend.A | |
| prison-punishment | re-naming | done | short-cut | state | misleading | |
| propagandistisch.A | | | | | | |
| propaganda | | | | | | |

When there are some shared nearest neighbours of the base vector and the derived vector (indicated by the bold face neighbours in (16)/(17)), the shift vector is basically noise and the meaning of the derived verb is compositional. That is, the combination of the verb *kleben* and the prefix *über* yields the APPLICATION meaning predicted by our lexical-conceptual classification, in which the meaning of the prefix and the derived verb is the same as the meaning of the preposition and the base verb in the locative alternation, see (6). In contrast, *schauen* ('to look')/*überschauen* ('to survey') as in (19)-(21) constitute a prototypical example where there are no salient shared neighbours of the base and the derived vector, but where the derived vector shares salient neighbours with the shift-vector.

(19) *schauen* (to look) BASE

| | | | | | | |
|---|---|---|---|---|---|---|
| gucken.V | starren.V | anstarren.V | anblicken.V | blicken.V | anschauen.V | angucken.V |
| peer | stare | at.PRTC.stare | look-at-so. | look | look-at-s.o. | peer-at-s.o. |
| grinsen.V | lächeln.V | reinschauen.V | | | | |
| grin | smile | look-into-s.th | | | | |

(20) *überschauen* (to survey) DERIVED

| | | | | |
|---|---|---|---|---|
| überblicken.V | **Komplexität.N** | **Tragweite.N** | Gestirn.N | Mannigfaltigkeit.N |
| survey | complexity | bearing | luminary | complexity |
| Einbildungskraft.N | Ansehung.N | **Gesamtzusammenhang.N** | Materie.N | |
| imagination | reputation | totality | interstellar-matter | |
| unüberschaubar.A | | | | |
| unmanagable | | | | |

(21) *über* (over) SHIFT

| | | | |
|---|---|---|---|
| **Komplexität.N** | Berücksichtigung.N | Folgewirkung.N | **Gesamtheit.N** |
| complexity | taking-into-account | consequence | totality |
| Verflechtung.N | Umwelteinwirkung.N | Beeinträchtigung.N | **Tragweite.N** |
| interconnection | environment-consequence | impairment | bearing |
| Funktionsträger.N | Differenzierung.N | | |
| administrator | differentiation | | |

We propose that when the overlap in nearest neighbours is greater between derived and shift vector ((20)/(21)) than between base and derived vector ((19)/(20)), this indicates that the meaning of the derived verb is figurative and that the meaning of the prefix *über* and the base verb *schauen* in combination is different from the meaning these words have in isolation. We call such a meaning of a complex expression that cannot be reduced to the meanings of its constituents "holistic". Tellingly, in contrast to *überkleben* (17), the base verb *schauen* is not among the nearest neighbours of the derived verb *überschauen* (20). The holistic semantic effect of prefixing *schauen* with *über* is linguistically reflected in the ungrammaticality of the locative alternation with *überschauen*. Whereas the meaning of the base verb *schauen* licenses the realization of the Ground argument with a PP-complement (22-a) but not as the direct object of a prefix-construction (22-b), the holistic meaning of the prefix verb *überschauen* licenses the Ground argument only as a direct object (23-b) but not as a PP complement (23-a).

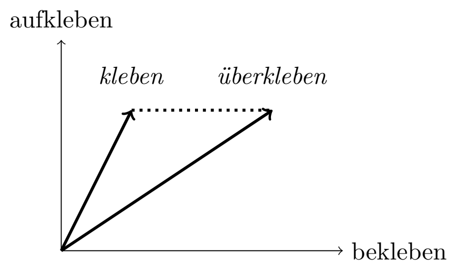
(22) a. Der Mann schaute über die Stadt.
the man look over.PREP the city
'The looked over the city.'
b. ?Der Mann überschaute die Stadt.
the man over-PRFX.see the city
'The man overlooked the city.'

(23) a. *Der Mann schaute über die Komplexität des Problems.
the man look over the complexity the.GEN problem
b. Der Mann überschaute die Komplexität des Problems.
the man over-PRFX.look the complexity the.GEN problem
'The man surveyed the complexity of the problem.'

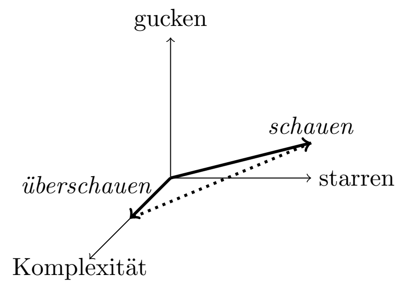
An intuitive explanation for the contrast between (22) and (23) may be given as follows. In (22), *schauen* is a perception verb that can be complemented with a PP specifying the perceptual space (i.e. that the subject has a view over the city). Consequently, because in (23) a spatial specification of the field of view with a PP is ungrammatical, this suggests that the relevant dimension of meaning in which *überschauen* is interpreted is no longer spatial, as would be expected for a verb that participates in the locative alternation. Instead, the composition of the verb and the prefix induces a holistic semantic effect by which the meaning of the prefix-verb is dislocated to a dimension of meaning not present in the prefix or the base verb in isolation. A quite similar holistic effect of meaning composition is involved in pure form in the fifth cluster of verbs of 'overpowering' (13), where the distributional characterization shows that the expected change of location reading is by and large replaced by the dislocated meaning of an unforeseeable event of (natural) force. In other words, whereas the meaning of the preposition and verb in the composition of *überkleben* is "rigid" (i.e. the meaning is not sensitive to context) and the salient dimensions of meaning of the preposition and the verb do not change through composition, the meaning of the preposition and verb in the composition of *überschauen* is 'non-rigid' and the salient dimensions of meaning of the preposition and the verb do change through composition. While such intuitions about the "dislocation" or "change" of a word's meaning dimensions are quite plausible when word meaning is perceived as a point in a high-dimensional vector space as DSM representations do, these intuitions are difficult to detect and represent in terms of lexical operations on the LCS of the base verb. Consequently, the way in which we phrased our intuitions hints towards the possibility that transparent DSM representations are better suited to make precise the semantic operation underlying the contrast between (22) and (23) on the one hand and *schauen* and *kleben* on the other. To foster an intuitive understanding of what it means that the meaning components denoted by the dimensions of a pair of vectors remain (mostly) unchanged in one case, but change in others, in the following we frame the contrast between (16)/(17) on the one hand and (22)/(23) on the other in a figurative understanding of meaning as a vector space. Thus, the following elaborations are neither intended as formally accurate explanations of DSM representations – in particular, we use nearest neighbours as approximations of dimensions – nor as lexical representations of word meaning in the traditional sense. Instead, we use the idea of meaning being represented in a vector space in a non-technical way to highlight what we believe is the specific "surplus" of DSM representations of meaning when compared against LCS-style analyses.
Consider first the simple rigid composition of *kleben* and *über*, where the base verb and the derived verb have salient nearest neighbours in common, i.e. the bold-faced nearest neighbours

in (16)/(17). For the sake of illustration, assume that we characterize the meaning of the base and derived verb with two of these shared salient nearest neighbours – (*bekleben* (paste sth. up ) and *aufkleben* (to glue sth. on)) – and interpret the vectors associated with these neighbours as the dimensions of the meaning of the base and derived verb. Second, in the holistic case (19)/(20), the derived verb and the shift vector but not the base and derived verb share salient dimensions of meaning. Assume for the sake of illustration that we characterize the base verb *schauen* with its two most salient nearest neighbours *gucken* ('to peer') and *starren* ('to stare') and the derived verb with its most salient nearest neighbour *Komplexität* and that we use the vectors associated with these nearest neighbours as the meaning dimensions of the base and derived verb. The figures (24) and (25) visualize the meaning spaces characterized by these assumptions, where we represent the contribution of *über* according to our additive composition model as a dotted vector.

(24) rigid meaning composition




(25) holistic meaning composition




In (24) the meaning components denoted by the dimensions of the vectors remain (mostly) unchanged, but are deleted or overwritten in (25). That is, in (24) the composition of *über* and the base verb retains the original meaning dimensions and adds new dimensions already present in the meaning of the base verb, but in (25) the meaning dimensions of the base verb are replaced with new ones not present in the meaning of the base verb. Figuratively speaking, the derived verb *überkleben* lives in the same meaning space in which the base verb lives. In contrast, *überschauen* lives in a region of the meaning space different from that in which the constituents *überschauen* is composed of are located. In sum, whereas rigid composition is dimension-preserving and the meanings of *über* and *kleben* are the meanings these words have in isolation, holistic composition is non-dimension-preserving and the meaning composed of *über* and *schauen* cannot be decomposed to the meanings the preposition and the base verb have in isolation. Concluding, what we intend to make tangible with (24)/(25) is that the relation between lexical-conceptual semantics and DSM representations is more complex than it appears at first glance. In particular, the differences between the two are not just of a technical but also of a conceptual nature; the high dimensionality of the meaning space encoded in a DSM captures aspects of verb meaning that cannot be detected and represented with lexical frameworks like LCS (which focus on specific meaning dimensions like event or argument structure). But precisely because the "surplus" of DSM representations of word meaning falls outside the scope of traditional lexical semantics, this raises the question for how phenomena like the holistic meaning composition in (25) can be operationalized in a way that is compatible with established frameworks of lexical semantics like LCS. Given these complimentary strengths of LCS and DSM models of word meaning, we believe that a further investigation of the combination of lexical-conceptual and usage-based approaches may lead to an empirically grounded and theoretically sound theory of word meaning in its entirety.

# 4 Conclusion and Outlook

By means of a case study, we aimed to show that transparent DSM representations, when compared with the more traditional approach of lexical-conceptual semantics, provide a novel and exciting way to investigate the conceptual underpinnings of verb meaning in an empirically grounded and theoretically unbiased way. However, throughout the paper we were at pains to limit our attention to the discussion of observations we made rather than attempting to put forward a systematic theory of DSM representations and the principles of their composition. We remained reluctant with respect to broad claims about the nature and status of DSM representations because we simply put aside a question which, although of fundamental importance, we were not able to address given the goals and limitations of this paper. While it is standardly assumed in the literature (without further argument) that DSMs represent the meaning of words, in our case study we assumed that DSMs represent conceptual features (in the sense of Levin's cohesive semantics features of Pinker's narrow range lexical rules) only loosely associated with a specific word. In order to develop a systematic theory of what it is that DSM representations encode and consequently how DSM representations figure in the view of compositional meaning computation advanced in formal semantics, we believe that it is necessary to get a better understanding of what the objects of meaning are that DSM representations encode, for it makes a difference whether we are concerned with a theory of concepts and their linguistic expression or a theory of linguistic expressions and their conceptual underpinnings.

## Bibliography

**[1]** M. Baroni and A. Lenci. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721, 2010.
**[2]** M. Baroni, R. Bernardi, and R. Zamparelli. Frege in space: A program of compositional distributional semantics. *LiLT*, 9:241 – 346, 2014.
**[3]** S. Clark. Vector space models of lexical meaning. In *The Handbook of Contemporary Semantic Theory*, pages 493 – 522. Wiley Blackwell, 2015.
**[4]** G. Faaß and K. Eckart. SdeWaC - a corpus of parsable sentences from the web. In *Proceedings of the International Conference of the GSCL*, 2013.
**[5]** A.Lenci. Carving verb classes from corpora. In *Word Classes: Nature, Typology and Representations*, p. 17 – 36. John Benjamins, 2014.
**[6]** B. Levin. *English verb classes and alternations: a preliminary investigation.* University of Chicago Press, 1993.
**[7]** B. Levin and S.Pinker. Introduction. In *Lexical & Conceptual Semantics*, p. 1–8. Blackwell, 1991.
**[8]** O. Levy and Y. Goldberg. Neural word embedding as implicit matrix factorization. In *Advances in Neural Information Processing Systems 27*, p. 2177–2185. Curran, 2014.
**[9]** T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Proc. of NIPS*, p. 3111–3119, 2013.
**[10]** S. Pinker. *Learnability and Cognition: The Acquisition of Argument Structure. New edition.* MIT Press, 2013.
**[11]** S. Schulte im Walde. Experiments on the Automatic Induction of German Semantic Verb Classes. *Computational Linguistics*, 32(2):159–194, 2006.
**[12]** P. D. Turney and P. Pantel. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37:141–188, 2010.
**[13]** F. Van der Leek. The english conative construction: A compositional account. In *CLS 32*, p. 363–378, 1996.