Integrating lexical-conceptual and distributional perspectives on word meaning

SFB 732 Final Colloquium.
Tillmann Pross (B4) (based on joint work with Antje Roßdeutscher (B4), Sebastian Padó (B9), Gabriella Lapesa (B9), Max Kisselew (B9))

1 Introduction

1.1 Two problems for a systematic theory of word meaning

- Problem of Analytical Depth

    - If *bachelor* means unmarried man, why is the pope not a bachelor?
    - 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?
    - Such particular definitions can be patched up, but sceptics foresee a never-ending need for such patching, with no real increase in watertightness.

- 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)

1.2 Two popular strategies to overcome the problems of depth and breadth

1.2.1 Lexical-conceptual semantics

- Hypothesis: "syntactic properties of phrases reflect, in large part, the meanings of the words that head them"(Levin and Pinker, 1991).

- Key observation: many verbs allow for so-called argument structure alternations, where one and the same verb can appear in more than one syntactic construction.

- For example, although there seems to be no obvious syntactic difference between the transitive uses of *to cut* and *to break*, these verbs differ with respect to participation in the conative alternation.

(1)   She broke the bread

(2)   She cut the bread.

(3)   *She broke at the bread.

(4)   She cut at the bread.

- To lexical-conceptual semantics, differences as in the conative alternation are not arbitrary but result from "systematic semantic and morphological differences between the verbs that enter into a construction and those that are syntactically similar but fail to enter into it" (Pinker, 2013).

- For the conative alternation, the relevant distinction has been argued to be of a conceptual nature: *cut* is a verb of motion, contact and causation whereas *break* is a verb of pure causation (Guerssel et al., 1985)

- These conceptual differences (a) are represented in the particular meaning of a verb in a way that syntax can be sensitive to and (b) when taken together, induce semantically cohesive verb classes.

1.3 Distributional Semantics

- Hypothesis: words that occur in similar contexts tend to have similar meanings, see Turney and Pantel (2010) for an overview.

- The distribution of a word's contexts are considered central to the construction of a suitable meaning representation of that word.

- A distributional 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 distributional 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*.

1.4 Even if the initial problems have been overcome, chronic untractabilities remain

- The conceptual underpinnings of verb meaning

  - One may well reproduce a given gold standard of classification with the methods of distributional semantics while still there is "little understanding of the meaning components, i.e. the semantic features, relevant to analyze verb meaning" (Lenci, 2014), see also Schulte im Walde (2006).

  - The assumption that contact and motion are required for a verb to enter the conative construction 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)

- Grey area of word meaning: semi-productive morphosemantics (vs. productive and transparent morphosemantics vs. unproductive and opaque morphosemantics)

  - "Even a very superficial look at a dictionary will deliver a pervasive picture of non-compositionality defying any quick generalizations" Kratzer (2003);

- Semi-productivity often correlates with primary data sparsity, but using morphological structure to overcome data sparsity (e.g. as in Padó et al. (2016)) requires transparent morphosemantics to fall back on.

- Main goal of B4: push back the boundaries of transparent analysis and see how far we can carry things. Empirically: prefix and particle verbs, nominalizations, prepositions; Conceptually: forces, qualities, non-final-non-initial-paths, ...

- Third phase of the SFB: Join forces with B9 to explore the conceptual underpinnings of verb meaning in the grey zone, using distributional models to sharpen our intutions and lexical-conceptual semantics as heel lifters.

---

2 The composition of meaning in the grey area

- Subject of study: German verbs prefixed with *über* (Pross et al., 2017).

- One of the vexing phenomena we have encountered is a semi-productive participation in the locative alternation.

- An example of fully productive participation in the locative alternation with *über* is given in (5)/(6): the prefix verb *überkleben* (to paste over) alternates with the verb *kleben* in which the direct object of the prefix verb is realized by a PP headed by *über*.

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

(6)   Peter klebte den Aufkleber über   den Kratzer.
      Peter paste the sticker   over.PREP the scratch.
      'Peter pasted the sticker over the scratch'

- With the verb *überschauen*, the locative alternation is semi-productive.

- The meaning of the base verb *schauen* licenses the realization of the Ground argument with a PP-complement (7) but not as the direct object of the figurative interpretation of the prefix-construction (8)

(7)   Der Mann schaute über   die Stadt.
      the man look   over.PREP the city
      'The looked over the city.'

(8)   ?Der Mann überschaute   die Stadt.
      the man   over-PRFX.see the city
      'The man overlooked the city.'

- Conversely, the meaning of the prefix verb *überschauen* licenses the Ground argument only as a direct object (10) but not as a PP complement (9)

(9)   *Der Mann schaute über die Komplexität des   Problems.
        the man   look   over the complexity   the.GEN problem

(10)   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.'

- What is it that prevents *schauen* from productive participation in the locative alternation?
- To gain deeper insights into this question, we adopted an additive model of the composition of distributional representations (see Baroni et al. (2014) for an overview) 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.
- The distributional representations we used were word embeddings with 300 dimensions learned with the Word2Vec algorithm from SdeWac (Faaß and Eckart, 2013) using a continuous bag of words model (Mikolov et al., 2013) with a symmetric 5-word window.
- Because these word embeddings are dense and thus non-interpretable by humans, we approximated the dimensions of the word embeddings with their nearest neighbours in the vector space (identifying proximity of two vectors with their dot-product (Levy and Goldberg, 2014))

2.1   Rigid composition

- Consider first the approximated distributional representations of the base verb *kleben* (to glue) (11), the derived verb *überkleben* (to cover) (12) and the shift vector in (13).

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

(12)   *überkleben* (to cover) DERIVED
         Aufkleber.N **bekleben.V**   Plakat.N Schriftzug.N Aufschrift.N **kleben.V aufkleben.V**
         sticker   be.PRXF.glue poster   letters   label   glue   on.PRTC.glue
         bedrucken.V Aufdruck.N prangen.V
         be-print   logo   be-respleshdent

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

- **Hypothesis**: when there are some shared nearest neighbours of the base vector and the derived vector (indicated by the bold face neighbours in (11)/(12)), the shift vector is basically noise and the meaning of the derived verb is compositional.

- When the salient dimensions of meaning of the preposition and the verb do not change through composition, we call the composition of the meanings of the preposition and verb "rigid" (in that the meaning of the complex construction is not sensitive to context)

2.2   Holistic composition

- Next, consider the approximated distributional representations of the base verb *schauen* (to look) (14), the derived verb *überschauen* (to survey) (15) and the shift vector in (16).

(14)   *schauen* (to look) BASE
gucken.V starren.V anstarren.V   anblicken.V blicken.V anschauen.V angucken.V grinsen.V
peer   stare   at.PRTC.stare look-at-so. look   look-at-s.o. peer-at-s.o. grin
lächeln.V reinschauen.V
smile   look-into-s.th

(15)   *überschauen* (to survey) DERIVED
überblicken.V **Komplexität.N Tragweite.N** Gestirn.N Mannigfaltigkeit.N Einbildungskraft.N
survey   complexity   bearing   luminary complexity   imagination
Ansehung.N **Gesamtzusammenhang.N** Materie.N   unüberschaubar.A
reputation   totality   interstellar-matter unmanagable

(16)   *über* (over) SHIFT
Komplexität.N Berücksichtigung.N Folgewirkung.N **Gesamtheit.N** Verflechtung.N
complexity   taking-into-account consequence   totality   interconnection
Umwelteinwirkung.N   Beeinträchtigung.N **Tragweite.N** Funktionsträger.N
environment-consequence impairment   bearing   administrator
Differenzierung.N
differentiation

- **Hypothesis**: when the overlap in nearest neighbours is greater between derived and shift vector ((15)/(16)) than between base and derived vector ((14)/(15)), the meaning of the prefix *über* and the base verb *schauen* in combination is different from the meaning these words have in isolation.
- Tellingly, in contrast to *überkleben* (12) where the base verb *kleben* is among the nearest neighbours of the derived verb *überkleben*, the base verb *schauen* (to look) is not among the nearest neighbours of the derived verb *überschauen* (to survey) (15)
- We call such a meaning of a complex expression that cannot be reduced to the meanings of its constituents "holistic".
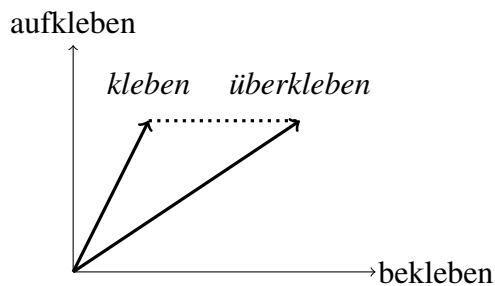
---

3   Intuitions meaning composition in vector space

- To foster an intuitive understanding of how distributional representations capture the fact that meaning components denoted by the dimensions of a pair of vectors remain (mostly) unchanged in one case, but change in others, we frame the contrast between rigid and holistic composition in a figurative understanding of meaning as a vector space.

- Consider first the 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 (11)/(12).
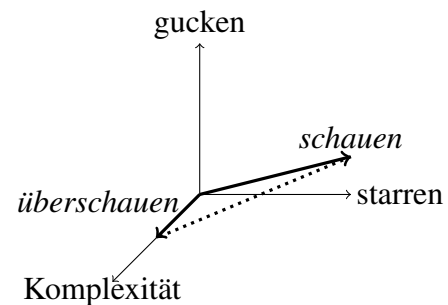
- For the sake of illustration, assume that we characterize the meaning of the base and derived verb with two of these shared salient nearest neighbours – (*bekleben* (to paste sth. up) and *aufkleben* (to glue sth. on)) – and interpret the vectors associated with these neighbours as the dimensions of the meaning of the base and derived verb.

- Second, in the holistic case (14)/(15), the derived verb and the shift vector but not the base and derived verb share salient dimensions of meaning.

- Again, 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* ('complexity') and that we use the vectors associated with these nearest neighbours as the meaning dimensions of the base and derived verb.

- The figures (17) and (18) 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.

(17) **rigid meaning composition**

(18) **holistic meaning composition**

- In (17) the meaning components denoted by the dimensions of the vectors remain (mostly) unchanged, but are deleted or overwritten in (18).

- In (17) the composition of *über* and the base verb retains the original meaning dimensions (i.e. new dimensions already present in the meaning of the base verb are added)

- In (18) 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.

6

4   Summary

- Returning to the question what restricts participation of *überschauen* in the locative alternation, the answer that the previous considerations on the interpretation of distributional models suggest is that productive participation in the locative alternation requires a rigid composition of the matrix verb but *überschauen* is composed holistically.

- Traditional view: the only mechanism of meaning construction accepted as valid is rigid composition

- This results in a distinction between compositional (=worthwile investigation) and non-compositional (=idiosyncratic, not worth investigation) meanings

- Our case study suggests that this view of meaning production is too simple and restricted.

- The integration of lexical-conceptual and distributional semantics forsters a more complex picture of the composition of (word) meaning than is traditionally assumed, where rigid and holistic composition are both systematic mechanisms of the production of meaning worth investigation.

- Of course, this raises the question for how phenomena holistic meaning composition can be operationalized in a way that meets the high formal standards of rigid semantics (and notably, the same question arises for the more well-researched case of adjective-noun composition, see e.g. Asher et al. (2016); McNally and Boleda (2017)).

- One approach to the formalization of holistic meaning composition we have been investigating recently (Pross (2018)) is the operation of 'reification', as understood in the work of Reichenbach (1947) and Davidson (1967): if you want to find out more about it, come to our poster!

References

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

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.

Donald Davidson. The logical form of action sentences. In Nicholas Rescher, editor, *The Logic and Decision of Action*, pages 81 – 95. The University of Pittsburgh Press, Pittsburgh, 1967.

Gertrud Faaß and Kerstin Eckart. SdeWaC - a corpus of parsable sentences from the web. In I. Gurevych, C. Biemann, and T. Zesch, editors, *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology*, 2013.

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.

Angelika Kratzer. The event argument. http://semanticsarchive.net/Archive/GU1NWM4Z/, 2003.

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 and Steven Pinker. Introduction. In *Lexical & Conceptual Semantics*, pages 1 – 8. Blackwell, 1991.

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.

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.

Sebastian Padó, Aurélie Herbelot, Max Kisselew, and Jan Šnajder. Predictability of distributional semantics in derivational word formation. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1285–1296. The COLING 2016 Organizing Committee, 2016. URL `http://aclanthology.coli.uni-saarland.de/pdf/C/C16/C16-1122.pdf`.

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

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. to appear.

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.

Hans Reichenbach. *Elements of Symbolic Logic*. The Macmillan Company, London, 1947.

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.