Towards a Computational Model of Gradience in Word Sense

Katrin Erk, University of Texas, Austin
Sebastian Padó, Saarland University, Saarbrücken

November 28, 2006

In the construction of lexicons or word sense inventories, researchers have usually strived to define a set of *disjoint* senses for a word; however, there is increasing evidence that this is impossible in the general case. In this programmatic paper, we support this point with *sense overlap* data from manual sense annotation. We then sketch the road towards a fundamentally *graded* representation of word sense that is still amenable to integration into a theory of sentence semantics.

Word Sense Overlap. In computational linguistics and traditional lexicography, it is usually assumed that the senses of a word form a set of distinct readings, and that any given instance evokes exactly one of these readings.

However, recent studies indicate that this assumption is a at least a simplification: senses often appear to *overlap*. For example, the annotators tagging the English SENSEVAL datasets could not reliably distinguish between the plant and food sense of "onion" [10]. In our own experience from German sense tagging [1], more than 3% of all verbal instances in a large corpus could not be assigned a single sense, with a much higher proportion for some predicates.

A simple example from our data is the German verb "überraschen" (*to surprise*), which has (at least) two senses which can be paraphrased as (A), an experiencer experiences an emotion of surprise, and (B), an experiencer is caught flat-footed by an event or agent.[1] Sentences (1) and (2) show clear instances of the individual senses; however, (3) shows an instance where annotators found it impossible to decide for a single sense, both being present to some degree.

(1) Ich wäre sehr **überrascht**, sollte Fellini nicht immer wieder mit dieser Frage konfrontiert worden sein. (TIGER s7058)
'I would be very **surprised** if Fellini wasn't often confronted with this question.'

(2) Mann im Schlaf von Steinschlag **überrascht**
'Man **surprised** by rock slide'

(3) [Die Wurzeln] mancher unguter Erscheinungen, die uns immer wieder **überraschen**, [liegen] in dem eben beschriebenen Zustand. (TIGER s11430)
'Some bad phenomena which **surprise** us originate in the described state.'

[1] A and B correspond to WordNet senses 1 and 2 of "surprise", respectively.

Another example is the German word "bemerken", which can mean either to notice something or to comment on something. There are numerous examples which seem to imply both, like the following:

(4) Kein Wunder, daß Gerhard Schäfer in seinem Buch derzeit eine "Renaissance der Verbindungen in den neuen Ländern" bemerkt. (TIGER s11777)
'(It is) not surprising that Gerhard Schäfer **notices/comments on** a "renaissance of fraternities in the new states".'

Word sense overlap shows the limits of treating word senses as a set of distinct readings. By undermining the appropriateness of classification techniques, it calls into question current computational models of word sense disambiguation. One response to this issue has been to suggest that the blame lies with current word sense inventories, and that overlap can be avoided by empirically discovering coarse-grained, distinct word senses [7]. However, we think that sense overlap cannot be explained away, and that it is in fact impossible to always identify distinct word senses. In fact, we can involve exactly those senses which show overlap in a zeugma effect, a classical linguistic test for the presence of two distinct senses:

(5) * Den Jungen überraschte das Gewitter im Wald, und dass es zum Frühstück Kakao gab.
'The boy was surprised by the thunderstorm in the forest, and that there was hot chocolate for breakfast.'

Similar observations have been made in related fields as well. The lexicographer Hanks [6], who has compared "bank" in the financial institution sense with "blood bank" and "data bank", found that the meaning of "bank" is best described using *meaning components* such as "storage" or "money", which can be present to different degrees. In a psycholinguistic study, Coleman and Kay have shown that the appropriateness of "lie" for certain situations can be predicted by using (potentially overlapping) properties of the situations [2]. These findings represent a body of evidence for graded word sense in search of an explanation.

Prototypes and Attributes. One theory that is centrally aimed at representing gradedness is *prototype theory* [16], which models membership in a concept as a *graded property*, defined as distance between the instance and the concept's prototype in a conceptual space.[2] While the origins of prototype theory lie in concept representation, it has been shown that the phenomena in this area show considerable parallelism to those in word sense [14], and prototypes have explicitly been suggested by Hanks for modelling word senses. Prototype models characterise concepts in terms of attributes, none of which need to be either necessary or sufficient [17]; an instance will be member of a concept to a higher degree if its attribute values are closer to those of the prototype (or the concept's exemplars, respectively).

Attribute-based models at first glance seem reminiscent of "traditional" feature-based theories of meaning such as [9], which have the problem of having to

[2]The rival *exemplar-based models* also predict graded membership, but define it as distance between the new instance and all remembered instances of the concept [12]. For modelling word sense, both seem equally well suited.

list meaning primitives that can be combined to provide a *full* account of word meaning. However, existing prototype models in lexical semantics [6, 2] typically do not attempt a full characterisation of word senses in terms of features, nor do they attempt to identify primitive features. Rather, they focus on finding (potentially non-primitive) attributes that explain the phenomena of meaning and meaning gradedness in the data.

In the case of our "überraschen" example above, we can explain both the distinctness of (A) and (B) and their overlap in terms of attributes: Assuming that (A) has the attribute "emotion of surprise", (B) has a attribute "being caught flat-footed", and both share the attribute "being taken by surprise", the difference in their attribute sets distinguishes the senses. We presume that both senses are perceived to be present when none of the distinguishing attributes of either (A) or (B) can be observed clearly, as is the case in (3).

Towards a Computational Model of Graded Word Sense. Our proposal is to use cognitive models, or more specifically prototype or exemplar models, also for word sense[3]. For such a model to be interesting for computational linguistics, two conditions have to be met: it must be possible (1), to integrate it with an account of sentence semantics; and (2), to construct it empirically.

As for (1), the model needs to be able to serve as part of a *compositional* account of semantics. That is, it needs to be possible to compute the meaning of a complex expression from the meanings of its components. It has been debated whether a prototype view of meaning is compatible with compositionality at all [15, 4]: The problem is that straightforward formalisations of prototype composition cannot model the possibility that exemplars are more prototypical for a composite concept ("striped apple" or "pet fish") than for its components. To address this problem, more complex mechanisms of composition have been proposed [8, 5]. In particular, [5] is interesting in that it provides an informal sketch of an attribute-based mechanism of composition, based on positive or negative correlation relations between attributes. We think that a formalisation of this mechanism is currently the most promising candidate for an effective compositional model of gradience in word sense.

In such an attribute-based model, (2) means primarily the practical determination of relevant attributes for each sense. Fortunately, a number of approaches to attribute acquisition have been proposed. Much work in psycholinguistics has been done on definition [18, 11] and association experiments [13]. One linguistic approach that requires manual labour, but reduces the reliance on a single lexicon designer's intuition, has been presented by Čulo [3], who induces attributes for semantic verb classes in German by a contrastive analysis of related classes.

Conclusion. In this paper, we have argued that gradience in word sense is a fundamental property which needs to be accounted for by any model of word

[3]Using prototype or exemplar models for lexical semantics presupposes a view of meanings as *conceptual* entities. While we consider this to be appropriate, it contrasts with the "realist" approach predominant in formal semantics [17], which sees meanings as entities in the real world.

sense, using sense overlap as an example. This goal can be achieved by adopting prototype or exemplar models for word sense. However, much work is still to be done until such models can provide a *computational* treatment of word sense, the first step being the formalisation of the notions involved.

References

[1] A. Burchardt, K. Erk, A. Frank, A. Kowalski, S. Pado, and M. Pinkal. The SALSA corpus: a German resource for lexical semantics. In *Proc. LREC*, 2006.

[2] L. Coleman and P. Kay. The English word "lie". *Linguistics*, 57, 1981.

[3] O. Čulo. Methoden zur manuellen und automatischen Verknüpfung semantischer Verbklassifikationen. Master's thesis, Saarland University, 2006.

[4] J. Fodor and E. Lepore. The red herring and the pet fish. *Cognition*, 58(2):253–270, 1996.

[5] J. Hampton. The combination of prototype concepts. In P. Schwanenflugel, editor, *The psychology of word meanings*. Lawrence Erlbaum Associates, 1991.

[6] P. Hanks. Do word meanings exist? *Computers and the Humanities*, 34(1–2), 2000.

[7] E.H. Hovy, M. Marcus, M. Palmer, S. Pradhan, L. Ramshaw, and R. Weischedel. OntoNotes: The 90% solution. In *Proc. HLT/NAACL*, 2006.

[8] H. Kamp and B. Partee. Prototype theory and compositionality. *Cognition*, 57:129–191, 1995.

[9] J.J. Katz and J.A. Fodor. The structure of a semantic theory. In J.J. Katz and J.A. Fodor, editors, *The Structure of Language*. Prentice-Hall, 1964.

[10] R. Krishnamurthy and D. Nicholls. Peeling an onion: the lexicographers' experience of manual sense-tagging. *Computers and the Humanities*, 34(1-2), 2000.

[11] K. McRae, G. S. Cree, M. S. Seidenberg, and C McNorgan. Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37:547–559, 2005.

[12] D. L. Medin and M. M. Schaffer. Context theory of classification learning. *Psychological Review*, 85:207–238, 1978.

[13] A. Melinger, S. Schulte im Walde, and A. Weber. Characterizing response types and revealing noun ambiguity in German association norms. In *Proc. EACL Workshop "Making Sense of Sense"*, Trento, Italy, 2006.

[14] G. Murphy. Meaning and concepts. In P. Schwanenflugel, editor, *The psychology of word meanings*. Lawrence Erlbaum Associates, 1991.

[15] D. Osherson and E. Smith. On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9(1):35–58, 1981.

[16] E. Rosch. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104:192–233, 1975.

[17] J. P. Sutcliffe. Concept, class, and category in the tradition of Aristotle. In I. Van Mechelen, J. Hampton, R. S. Michalski, and P. Theuns, editors, *Categories and Concepts: Theoretical Views and Inductive Data Analysis*. Academic Press, 1992.

[18] D. P. Vinson and G. Vigliocco. A semantic analysis of noun-verb dissociations in aphasia. *Journal of Neurolinguistics*, 15:317–351, 2002.