

Can distributional approaches improve on Good Old-Fashioned Lexical Semantics?

Ann Copestake

Computer Laboratory, University of Cambridge
aac10@cl.cam.ac.uk

Abstract

In this position paper, I discuss some linguistic problems that computational work on lexical semantics has attempted to address in the past and the implications for alternative models which incorporate distributional information. I concentrate in particular on phenomena involving count/mass distinctions, where older approaches attempted to use lexical semantics in their models of syntax. I outline methods by which the earlier models allowed the transmission of information between lexical items (regular polysemy and inheritance) and address the possibility that similar techniques could usefully be incorporated into distributional models.

1 Introduction

While there has been much recent discussion of techniques for developing compositional approaches to distributional semantics, especially with respect to particular categories of phrase (e.g., adjective-noun), as far as I am aware, there has been no attempt to discuss systematically all the roles that distributional semantic representations might play in the production of a model of a sentence. Indeed, from the viewpoint of researchers working on ‘traditional’ areas of computational linguistics, such as parsing and generation, and those primarily interested in modeling language for its own sake, rather than application-building, the extensive work on distributional semantics has been somewhat disappointing in failing to provide models which are integrated with existing work to help solve long-standing problems. In some respects, most work on distributional semantics lacks ambition compared to earlier research on lexical semantics, in that previous approaches at least attempted to provide accounts that were fully integrated with syntax and full-coverage compositional semantics: i.e., which used lexical semantics as part of the models that assigned syntactic structure or logical form.¹ There are reasons to think that distributional approaches could well be more appropriate in such contexts, but a demonstration of this will involve looking at a broad range of phenomena. This paper is intended as a first step in outlining some of the issues that might be considered.

I first want to distinguish the discussion of lexical meaning here from the various approaches to deriving distributional meaning from sentences investigated by Clark and Pulman (2007), Baroni and Zamparelli (2010), Mitchell and Lapata (2010), Guevara (2011) and others which in turn relates to previous approaches to combining connectionist and symbolic approaches (e.g., Smolensky and Legendre, 2006). That line of work assumes that a syntactic representation (or perhaps a logical form) is available to guide the process of composition of distributions.² This work is mostly orthogonal to the issue I wish

¹Note that use ‘compositional semantics’ in its predominant sense to mean an approach in the tradition of Montague grammar, construed broadly, but including a treatment of quantification.

²I note the possibility of working with logical forms, since, although it is usual to work with syntactic relationships when working on compositional distributional semantics, the assumption is that these relationships are semantically meaningful. It thus seems possible that the models would, in principle, perform better if they were built on the basis of a logical form of some type, in that this provides a level of abstraction with respect to (some) verb alternations, expletive subjects and so on. Logical forms generally reflect a ‘deeper’ analysis which incorporates semantics associated with constructions, such as compound

to discuss here, which is whether the lexical phenomena addressed by earlier approaches might be modelled distributionally and whether this has implications for the overall architecture: for instance, in those case where lexical semantics affects syntax, some mechanism is required in the overall architecture to make syntax sensitive to the lexical semantic representation. This is not to say that there are no points of contact. For instance, in the notion of cocomposition described in Pustejovsky's Generative Lexicon (GL) work (e.g., Pustejovsky, 1995) the composition function is determined both by functor and argument. This can be perhaps related to some of the more recent work on composition with distributional semantics, where individual words can be associated with different composition functions (as suggested by Washtell (2011)). But GL is an exception in treating composition as part of a theory of lexical semantics, and even GL makes rather conventional assumptions about compositional semantics in many respects. Hence discussion of this is not part of the current paper.

I will concentrate here on research on modelling the behaviour of individual words rather than work on the traditional relationships between words (or word senses) — hyponymy, synonymy, antonymy and meronymy. Though this is not the focus of the current discussion, I will briefly touch on the use of hyponymy relationships in modelling the semantics of individual lexemes in §4.

At this point, a nomenclature issue arises, since there is no good collective term for the non-distributional approaches. 'Non-distributional' is clunky. To talk about 'traditional' or 'classical' lexical semantics seems inappropriate given the the earliest distributional work (e.g., Harris, 1954) predates, for example, the feature-based approach of Fodor and Katz (1963) (the first computational work on distributions was underway at this point, although the first publication I am aware of is Harper (1965)). The term 'symbolic' is problematic, since distributional semantics is also symbolic. So, in the absence of a better alternative, I will use 'Good old-fashioned lexical semantics' (GOFLS) by analogy with Haugeland's 'Good old-fashioned AI' (GOFAI: Haugeland, 1985). Hence the question that forms the title of this paper: "Can distributional approaches improve on Good Old-Fashioned Lexical Semantics?"

Models using hand-crafted GOFLS were integrated into parsing in a range of approaches from the 1970s onwards. For example, Boguraev (1979) used semantic preferences expressed in terms of semantic primitives specified by Wilks (1975) for disambiguation with an augmented transition network (ATN) parser. More complex models were later investigated within feature structure formalisms, perhaps most extensively within Pustejovsky's Generative Lexicon (GL) framework (Pustejovsky, 1995). Such approaches combine syntax, compositional and lexical semantics within one model and thus lexical semantics can influence and constrain syntax. This type of approach had some success in the 1980s and early 1990s in limited domains, but failed to scale to broad-coverage NLP. However, the models were (and are) nevertheless of interest to linguists and to psycholinguists. Seen from the perspective of using computational modeling to formally investigate language, they have therefore been partially successful.

Nevertheless, I think it is plausible to claim that the failure of GOFLS approaches in a computational setting was not just due to lack of resources to build highly complex lexicons, but to underlying problems with models that do not cope well with the 'messiness' of the actual data. Verspoor's detailed corpus investigation of some of the 'classic' GL cocomposition phenomena (Verspoor, 1997) is a case in point: to allow for the data there with a GOFLS model would have required fine-grained distinctions to be drawn which were otherwise unmotivated. Since that was precisely the problem with previous approaches to lexical semantics that had partly motivated the development of GL (see Pustejovsky's discussion and criticism of sense enumeration, for example), there was reason to doubt the classic GL model on theoretical grounds. Distributional-style approaches have been successfully adopted as models in investigation of some of the 'classic' GL phenomena (e.g., Lapata and Lascarides, 2003). However, these models are partial in that the distributional techniques have been used in isolation, rather than as part of an integrated syntactic-logical-distributional model. Furthermore, the aim in most published work is to show the best performance on a particular test set, rather than to build models which demonstrate good performance on a broad range of phenomena, let alone build fully-integrated broad-coverage systems.³

nouns and NPs acting as temporal modifiers. These might be expected to be relevant to the choice of functions for combining distributions.

³Baroni and Lenci (2010) argue convincingly that researchers should look at the performance of a single distributional

It therefore seems worthwhile to revisit some of the roles that GOFLS played in the earlier work, to investigate whether distributional semantics is really a promising alternative and to look at the requirements for distributional models under these assumptions. The viewpoint here is a theoretical/formal one (rather than practically-oriented NLP): what role can distributional models play in accounts of lexical meaning that aim to be linguistically (and psycholinguistically) plausible? The current paper is very preliminary — it concentrates on issues relating to the interaction of syntax and lexical semantics with respect to the count/mass distinction, and on the treatment of regular polysemy.

I will draw a distinction between the use of distributional techniques for acquisition of lexical semantic information for a GOFLS approach and models which use distributions directly. For instance, some approaches to interpreting compound nouns use semantic primitives to represent the relationships between the elements in the compound (such as Levi's classes: BE, HAVE and so on (Levi, 1978)). If these classes form part of the representation for the utterance, or are used in other processing, then even if the classes are determined via distributions, the final model is non-distributional. In contrast, a genuinely distributional model would represent the relationships themselves as distributions. Of course, the status of the primitives is not always clear in particular experiments: they may be seen as a convenient way of categorizing classes of distributions, for instance for evaluation purposes. Without the integration of models into larger frameworks, such distinctions are naturally a little fuzzy.

One deliberate omission here is any discussion of disambiguation or selectional preferences. It seems very plausible that distributions might be used to improve a parse-ranking model, and it is surprising there has been so little published work in this area, since it would seem a very useful way of evaluating different distributional techniques. That is, I would expect a good distributional model to be able to capture the sort of information about semantics that is necessary to resolve some proportion of coordination and PP-attachment ambiguities, and to be a much more satisfactory way of doing this than the earlier semantic primitive approaches. However, disambiguation in principle requires open-ended models of concepts. That is, in order to disambiguate some utterances, detailed knowledge of the world is required (as has long been recognised e.g., Fodor and Katz (1963)). To take a specific example:

- (1) Follow the path from the bend in the road to the car park.

It is reasonable that distributional semantics might allow partial disambiguation of the PP-attachment (e.g., determining that 'in the road' attaches to 'bend'), but without context (which might only be apparent on the ground rather than in the text) it is not clear how to attach 'to the car park'. Indeed, examples of this type often cannot be disambiguated by human annotators who lack access to the full context. For this reason, we cannot use disambiguation examples to test what information needs to be accessible in principle in a particular model, since in the worst case any information could be relevant (i.e., disambiguation is AI-complete).

In this paper, I will use two interrelated phenomena in order to look at how distributional semantics might replace GOFLS and what sort of models might be required. In §2, I will discuss some semantic constraints on grammatical behaviour. A variety of phenomena related to regular polysemy are then discussed in §3.

2 Distributional semantics and syntactic distinctions

There are a number of roles that lexical semantics could/should play in a grammar. Perhaps the most fundamental is to ensure that constraints on syntactic behaviour that relate to semantic categories can be represented and that constraints on the relationship between syntactic behaviour and meaning can be captured.

For example, in English, uses of nouns which denote humans in an utterance may not be mass terms.⁴ For example, (2) and (3) are ungrammatical because human-denoting nouns may not take *much*

model in a very broad range of contexts, but few published papers do this.

⁴This is an oversimplification. A full statement requires discussion of some of the complications of the mass/count distinction. For instance, there are nouns such as *troops* and *police* which are not classical count nouns because they have idiosyncratic

as a determiner.

(2) * Much children hate cabbage.

(3) * Much crowd was on the street.

An account of this generalization in a GOFLS framework might, for instance, state that lexical entries for all human-denoting noun lexemes inherit from a single general class, which has the desired syntactic properties associated with it. Numerous ways of implementing such generalizations have been developed, some incorporating defaults in the formalism so that exceptions could be allowed for. In any such approach, it is important that the semantic class can be justified and that multiple properties are predicted. For instance, the human-denoting nouns could also be predicted to occur with the relative pronoun *who* rather than *which*.

This assumes a lexicalist view of syntax. Some linguists (e.g., Borer, 2005) have argued that lexical entries do not specify detailed subcategorization information, mass/count distinctions and so on. The fact that grammaticality judgments involving subcategorization are graded rather than absolute can be taken to support such a view. Borer's approach is unimplemented (with the exception of Haugereid (2009)) but her viewpoint can, in fact, be seen as consistent with the way that Penn TreeBank derived grammars behave, in not ruling out utterances such as (2) or (3) or examples which violate subcategorization constraints (e.g., (4)).

(4) * I enjoy to run.

Indeed, even the broad coverage English Resource Grammar (ERG, Flickinger, 2000), which adopts an approach to syntax based on HPSG, and has a detailed lexicalist account of subcategorization which blocks examples such as (4), leaves most nouns as underspecified for count / mass distinctions, because so many nouns can appear in either mass or count contexts.

In a lexicalist account, if a lexeme like *lawyer* is marked as count, an utterance such as (5) is typically treated as ungrammatical (or extra-grammatical).

(5) In our legal method there is too much lawyer and too little law. [G. K. Chesterton]

It could only be interpreted by creating an extended (mass, non-human) use of *lawyer* (e.g., via lexical rule, in the manner discussed in the next section).⁵ In a construction-based account, such as Borer's, this use of *lawyer* simply ends up as being marked as mass. In and of itself, this does not indicate that the sentence is in any way odd, or that the meaning differs from the count use of *lawyer*. A GOFLS account could perhaps be combined with a construction-based approach to enforce the constraint on human-denoting terms in a way which would result in *lawyer* being marked as non-human-denoting in (5), though, as far as I am aware, such an account has not been proposed in detail, let alone implemented.

The theoretical disadvantage of GOFLS combined with a lexicalist approach is that it requires additional mechanisms to account for examples such as (5) and constraining such mechanisms is difficult. The approach is often criticized as being over-stipulative. In contrast, the disadvantage of GOFLS combined with the constructional account would be that there is no indication that examples such as (5) are in any way odd or rare. Conversely, there is the problem that mass readings are available in contexts which are underspecified for mass/count, such as (6).

(6) The lawyer came into the room.

behaviour with numerals.

⁵Example 5 could be taken to be metalinguistic, but it is reasonably representative of the sort of examples cited in the linguistics literature to show that all lexemes have both count and mass uses. In (5), I would take *lawyer* to refer to a property rather than being human-denoting (in the sense of referring to an individual or groups of individuals). In very general terms, this meaning shift is predictable, in that it is one of a range of possible types of use of predominantly count nouns in a mass context, but it is not the sort of use that would be listed by a lexicographer, for instance. So at least in that respect, it is distinct from the cases of regular polysemy, discussed in §3.

Intuitively, at least, this seems wrong: mass uses of predominantly count nouns should only be available in marked contexts.

We can sketch an alternative distributional account which begins to address such problems. For current purposes, I will just describe how a distributional approach might be integrated with a construction-based grammar. The first thing to note is that any such account requires partitioning or clustering the distributional space for the nouns. The constraint that a human-denoting term cannot be mass is assumed to apply to uses, rather than to words/lexemes.⁶ Nouns such as *lawyer* will be overwhelmingly count rather than mass, but the construction account allows for possibilities such as (5). For the time being, let's assume that the non-count/property use of *lawyer* is attested in the contexts from which the distributional model has been constructed (a possibly implausible assumption which I will return to below in §3). Of course the usual count use of *lawyer* will be much more frequent. If the contexts for *lawyer* include the determiners associated with it, the use of *much* will (hypothetically) only occur with a small numbers of uses. If the space of uses is partitioned or clustered into human-denoting vs property-denoting, contexts with *much* should only occur with the property uses. The boundary between human-denoting and non-human-denoting uses will be fuzzy, of course.

For the correlation with syntax to work, it must also be possible to partition the space of uses according to the count/mass behaviour. Clearly, whether a noun occurs with *much* would be directly accessible from a conventional distribution (if determiners were included), but other reflexes of count/mass behaviour require an extended notion of distribution, allowing sensitivity to morphological marking or plurality. It would be inappropriate to go into a detailed discussion of syntax here, so I will assume for simplicity that the count/mass distinction is binary, that all instances of a noun in an utterance can be marked as count, mass or underspecified, and that contexts contain such information. If the constraint that human-denoting noun uses are never mass terms is valid, then we would expect the human-denoting space in a distribution to only contain uses marked as count or underspecified. The generalization that human-denoting terms are never mass could (at least potentially) arise from distributions of the relevant nouns rather than being stipulated.

The only piece of work which I am aware of which looks at count-mass distinctions using distributions is Katz and Zamparelli (2011).⁷ The paper demonstrates an initial result, which suggests that nouns which show large differences in semantics between singular and plural forms as measured using distributional techniques are predominantly mass (in that they are frequently found in contexts which select for mass terms, and infrequently found in contexts that select for count terms). This would fit with the assumption that some sort of meaning shift has to occur for a mass noun to be pluralized. However, the use of distributions here is limited to measuring semantic (dis)similarity. Building more complex models would require a corpus which makes distinctions between count and mass contexts systematically. The ERG-parsed Wikiwoods corpus (Flickinger et al., 2010) contains such information, but it is unclear whether this is sufficiently accurate to allow the relevant meaning shifts to be detected.

So this outline suggests something about the types of models that are of interest. Distributions must be sensitive to distinctions such as count / mass. If we take this as a syntactic distinction, then the appropriate models are ones in which distributions contain syntactic information.⁸ The advantage of the distributional model over the GOFLS approaches is that frequency effects are an integral part, and hence there is a natural account of the oddness of examples such as (5). The problem, from a practical perspective, is that distributions created over individual instances produce a severe sparse data problem (cf Rapp, 2004).

It is also, of course, implausible to assume that unusual cases such as that illustrated in (5) will actually be attested for all lexemes where they are possible in principle. What is actually required is an approach where certain uses may be postulated even though not actually attested with a particular word. Rather than discuss this with respect to marginal examples such as (5), I will turn to the phenomenon of

⁶Of course, the distributions for mass and count versions of a lexeme could just be constructed separately, but this is analogous to the simplistic lexicalist account where there are multiple, unrelated, word senses.

⁷I am grateful to an anonymous reviewer for drawing my attention to this paper.

⁸Though, in fact, there are arguments in favour of treating count / mass as part of compositional semantics, for instance by having sorts on variables which distinguish between divisible and indivisible.

regular polysemy.

3 Regular polysemy

The term regular polysemy is used to refer to the phenomenon that word senses (or usages) are often related to one another and that similar patterns of senses are found in groups of words. For instance, in most cases the same word is used for animals and their meat, (e.g., *lamb*, *turkey*, *haddock*, but not *deer/venison*) with the animal use being count and the meat use mass. This can be seen as a sub-case of a general pattern of count-mass conversions, which has been generically referred to as ‘grinding’. Regular polysemy has been extensively investigated in GOFLS. The empirical motivation for these accounts came from lexicography, and some of the computational implementations made use of information extracted from machine readable versions of conventional dictionaries (MRDs).

In an utterance such as:

(7) I’ve never seen so much turkey.

turkey is taken to be non-count. The role of lexical semantics is to ensure that this is associated with the correct meaning of the term (i.e., the meat rather than the animal sense). It should also ensure that a similar correlation can be made even in the case where the mass usage is previously unseen. For instance, speakers can understand a use of *crocodile* as in (8) and also generate it in an appropriate context, even if *crocodile* has not been seen as a mass term previously.

(8) I’ve never seen so much crocodile.

In a lexicalist account which associates mass/count with lexical entries, a new lexical entry for *crocodile* can be generated via a lexical rule, if *crocodile* is known to be of the appropriate type (e.g., ‘animal’). See, for instance, Figure 1, taken from Copestake and Briscoe (1995). The full details of the rule encoding are irrelevant here, but the following points should be noted. ‘1’ indicates the specification of the input to the lexical rule (the count term) and ‘0’ the output. The boxed integers indicate information sharing, so, for instance, the rule does not affect spelling (‘ORTH’) because the input and output share the same value. GL ‘qualia structure’ is used to represent aspects of lexical semantics. The compositional semantic representation is to be interpreted as producing a new predicate from the input (e.g., if the input semantics were equivalent to $\lambda x[\text{rabbit}(x)]$ the output would be $\lambda x[\text{grinding}(\text{rabbit})(x)]$). The syntactic effects come from the overall type of the structure (**lex-count-noun** and **lex-uncount-noun**). Lexical rules of this type can also be used for derivational morphology, which is relevant because some derivations show semantic relationships very similar or even identical to regular polysemy patterns.

An alternative approach (e.g., in Pustejovsky, 1995) involves combining the different senses/usages in a single structure via ‘dot objects’ (e.g., ANIMAL • MEAT). The assumption is that there are some regularities in the combination of types which are possible. Some contexts will select the ANIMAL use of a lexical item, while others will select the MEAT use. The dot object approach allows the ambiguity between the uses to be retained in some utterances, unlike the lexical rule account, but it is unclear whether this actually agrees with the linguistic and psycholinguistic evidence for this class of examples.

There are a number of criticisms that have been leveled at these different accounts, which I will not attempt to summarize here. Both, however, allow for regular polysemy as a fact about language which is to some extent conventional, rather than a fact about the world. This is much clearer with regular polysemy than with the marginal examples such as (5), since different languages show different polysemy patterns, and meaning shifts corresponding to regular polysemy in one language may be marked syntactically or by derivational morphology in others.

Regular polysemy has not been investigated much within distributional semantics (although see Boleda et al., 2012). Again, if there is a syntactic reflex, it is necessary to have a model which integrates this with the distributions to fully capture the effects. However, the point I want to discuss here is whether patterns in distributions can be used to predict semantic spaces which are too rare to be seen in the distributions of some lexemes. I take it that this reflects the situation which a human is in who

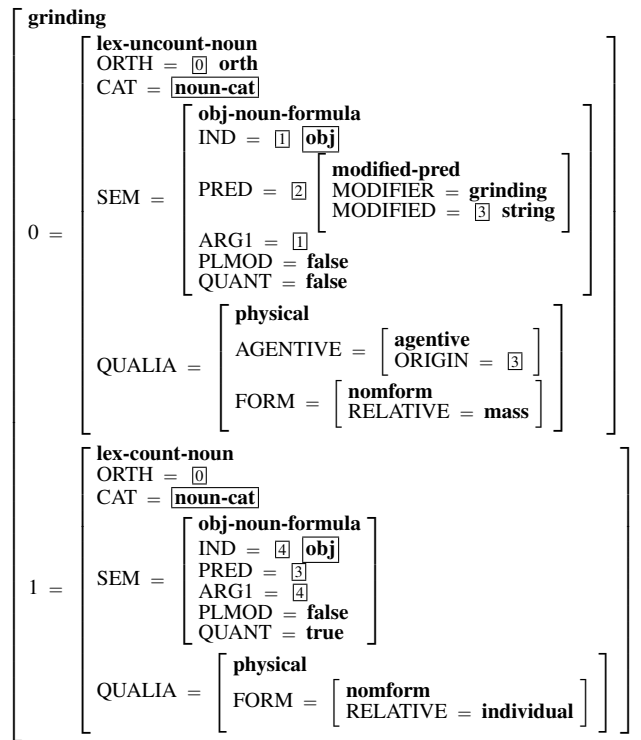


Figure 1: Grinding lexical rule from Copestake and Briscoe (1995)

hears an example such as (8) having never heard *crocodile* used as a mass term before. Schematically, we can imagine that the semantic spaces for words are as shown in Figure 2, where the unfilled circle by *crocodile* is supposed to indicate that this is a use that could be predicted based on the polysemy observed for other words, even though that use has not been observed by that particular hearer.

The theoretical attraction of such an account is that it incorporates frequency effects. It is neutral as to whether the different usages are to be taken as different senses: what it requires is just that the space of usages be partitionable. Whether novel uses could actually be predicted in this way is an empirical question, of course.

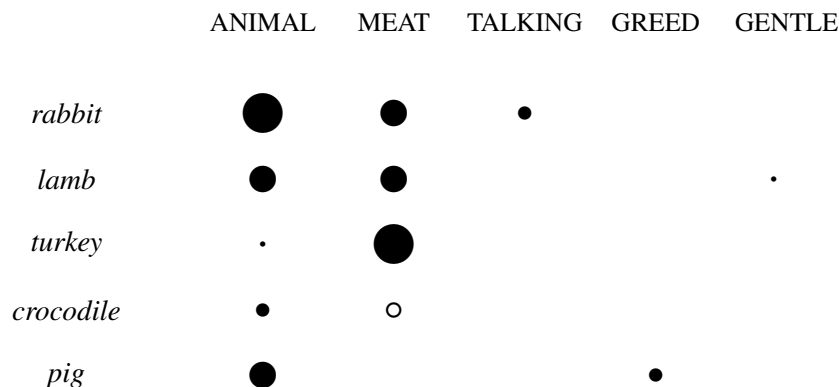


Figure 2: Schematic description of regular polysemy in terms of distributional spaces

4 Inheritance structure

A notable distinction between GOFLS accounts and distributional approaches is that most GOFLS approaches rely more-or-less heavily on some form of hierarchical structuring. In computational accounts, this can be used to allow inheritance or default inheritance. For instance, the GL qualia structure associated with the lexeme *book* might be inherited by *novel*. This allows semi-automatic construction of lexical entries with detailed lexical semantic information: for instance, in some earlier work taxonomies derived from MRDs were used to provide inheritance hierarchies and information about roles manually stipulated for the upper nodes only.

There is, of course, extensive computational work on deriving ontological relationships from corpora which is distributional in a broad use of the term, and also work on deriving such relationships from distributions in the narrower sense (e.g., Baroni et al., 2008). However, distributional models do not make use of inheritance relationships between words. Contexts which express hyponymy relationships, such as (9), will result in distributions for the hyponym which contain the hypernym (and vice versa), but that dimension is not distinguished in any way in the standard distributional approaches.

(9) Geese are waterfowl belonging to the tribe Anserini of the family Anatidae.

One way of thinking about the role of inheritance in GOFLS models is as a way of supplementing information about individual lexical items. For instance, if information about the qualia structure of a particular lexeme cannot be directly acquired, it might be obtained via inheritance. In an analogous manner, there seems to be scope for using automatically acquired ontological information in conjunction with distributional models, in particular to enrich the models of less frequent words. Distributional models require a considerable number of instances of words for good performance (and thus rely on the use of corpora which are vastly greater in size than anything which could plausibly correspond to the experience of an individual language learner). Ontology extraction systems, in contrast, achieve good performance on extraction of IS-A relationships with a single instance, provided the context for that instance is definitional in nature (dictionary definitions, Wikipedia articles and so on). It would thus seem natural to attempt to combine the two.

5 Conclusions

What I hope to have illustrated in this paper is that, to replace GOFLS accounts, distributional approaches will have to interact with syntax in a more integrated way than they currently do. That is, it is not enough to assume that distributions are created from syntactically parsed corpora and that distributions are composed in a manner guided by syntax, but that additionally syntax would have to be affected by distributions. I have tried to discuss ways in which distributional models could improve on GOFLS, and to suggest that they could, in fact, form part of the solution to some current linguistic debates.

The fact that distributional models are derived automatically from corpora is obviously a very strong point in their favour. But GOFLS models constructed from MRDs had an empirical basis too, and indeed, with the more modern dictionaries, the data was to some extent derived from corpora, albeit mediated by lexicographers. While there are obviously practical reasons to try and acquire all data directly from corpora, and while this makes the approaches more psycholinguistically plausible (if plausible corpora are used), there may nevertheless be ways in which more definitional information could and should also be incorporated. For instance, I have suggested above that there may be a role for models which use corpus-derived ontological relationships to supplement the usual derivational models.

The topics I have outlined here are just a small selection of those which could have been discussed: taken as a whole I believe the comparison with prior work suggests the need for some more ambitious theoretical work on distributional approaches that takes into account more of the linguistic issues that have driven past work on lexical semantics.

Acknowledgements

This paper arises out of joint work with Aurélie Herbelot (currently unpublished, but available from <http://www.cl.cam.ac.uk/~aac10/papers/lc-current.pdf>). I am grateful to her and to the anonymous reviewers for comments. All errors (of commission and omission) are my own.

References

- Baroni, M., S. Evert, and A. Lenci (2008). ESSLLI 2008 workshop on distributional lexical semantics.
- Baroni, M. and A. Lenci (2010). Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics* 36(4), 673–721.
- Baroni, M. and R. Zamparelli (2010). Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP10)*, pp. 1183–1193.
- Boguraev, B. (1979). *Automatic resolution of linguistic ambiguities*. Ph. D. thesis, University of Cambridge.
- Boleda, G., S. Padó, and J. Utt (2012). Regular polysemy: A distributional model. In *Proceedings of *SEM*, pp. 151–160.
- Borer, H. (2005). *Structuring Sense*. Oxford University Press.
- Clark, S. and S. Pulman (2007). Combining Symbolic and Distributional Models of Meaning. In *Proceedings of the AAAI Spring Symposium on Quantum Interaction*, Stanford, CA, pp. 52–55.
- Copestake, A. and T. Briscoe (1995). Semi-productive polysemy and sense extension. *Journal of Semantics* 12:1, 15–67.
- Flickinger, D. (2000). On building a more efficient grammar by exploiting types. *Natural Language Engineering* 6(1), 15–28.
- Flickinger, D., S. Oepen, and G. Ytrestøl (2010). Wikiwoods: Syntacto-semantic annotation for english wikipedia. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pp. 1665–1671.
- Fodor, J. and J. Katz (1963). The structure of a semantic theory. *Language* 39(2), 170–210.
- Guevara, E. (2011). Computing semantic compositionality in distributional semantics. In *Proceedings of the Ninth International Conference on Computational Semantics (IWCS 2011)*, Oxford, England, UK, pp. 135–144.
- Harper, K. E. (1965). Measurement of similarity between nouns. In *Proceedings of the 1st International Conference on Computational Linguistics (COLING65)*, New York, NY, pp. 1–23.
- Harris, Z. (1954). Distributional Structure. *Word* 10(2-3), 146–162.
- Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. MIT Press.
- Haugereid, P. (2009). *A constructionalist grammar design, exemplified with Norwegian and English*. Ph. D. thesis, NTNU, Norwegian University of Science and Technology.
- Katz, G. and R. Zamparelli (2011). Quantifying count/mass elasticity. In *Proceedings of 29th West Coast Conference on Formal Linguistics*.

- Lapata, M. and A. Lascarides (2003). A Probabilistic Account of Logical Metonymy. *Computational Linguistics* 29(2), 261–315.
- Levi, J. (1978). *The syntax and semantics of complex nominals*. Academic Press New York.
- Mitchell, J. and M. Lapata (2010). Composition in Distributional Models of Semantics. *Cognitive Science* 34(8), 1388–1429.
- Pustejovsky, J. (1995). *The Generative Lexicon*. MIT Press.
- Rapp, R. (2004). A practical solution to the problem of automatic word sense induction. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, pp. 26. Association for Computational Linguistics.
- Smolensky, P. and G. Legendre (2006). *The Harmonic Mind*. MIT Press Cambridge, MA.
- Verspoor, C. (1997). *Contextually-dependent lexical semantics*. Ph. D. thesis, University of Edinburgh. School of Informatics.
- Washtell, J. (2011). Compositional expectation: A purely distributional model of compositional semantics. In *Proceedings of the Ninth International Conference on Computational Semantics (IWCS 2011)*, pp. 285–294.
- Wilks, Y. (1975). A preferential, pattern-seeking, semantics for natural language inference. *Artificial Intelligence* 6(1), 53–74.