Distributional Semantics in Use

Raffaella Bernardi[*] **Gemma Boleda**[*] **Raquel Fernández**[†] **Denis Paperno**[*]
[*]Center for Mind/Brain Sciences
University of Trento
[†]Institute for Logic, Language and Computation
University of Amsterdam

Abstract

In this position paper we argue that an adequate semantic model must account for language in use, taking into account how discourse context affects the meaning of words and larger linguistic units. Distributional semantic models are very attractive models of meaning mainly because they capture conceptual aspects and are automatically induced from natural language data. However, they need to be extended in order to account for language use in a discourse or dialogue context. We discuss phenomena that the new generation of distributional semantic models should capture, and propose concrete tasks on which they could be tested.

1   Introduction

Distributional semantics has revolutionised computational semantics by representing the meaning of linguistic expressions as vectors that capture their co-occurrence patterns in large corpora (Turney et al., 2010; Erk, 2012). This strategy has been shown to be very successful for modelling word meaning, and it has recently been expanded to capture the meaning of phrases and even sentences in a compositional fashion (Baroni and Zamparelli, 2010; Mitchell and Lapata, 2010; Grefenstette and Sadrzadeh, 2011; Socher et al., 2012). Distributional semantic models are often presented as a robust alternative to representing meaning, compared to symbolic and logic-based approaches in formal semantics, thanks to their flexible representations and their data-driven nature. However, current models fail to account for aspects of meaning that are central in formal semantics, such as the relation between linguistic expressions and their referents or the truth conditions of sentences. In this position paper we focus on one of the main limitations of current distributional approaches, namely,

their unawareness of the unfolding discourse context.

Standardly, distributional models are constructed from large amounts of data in batch mode by aggregating information into a vector that synthesises the general distributional meaning of an expression. Some of the recent distributional models account for contextual effects within the scope of a phrase or a sentence, (e.g., (Baroni and Zamparelli, 2010; Erk et al., 2013)), but they are not intended to capture how the meaning depends on the incrementally built discourse context where an expression is used. Since words and sentences are not used in isolation but are typically part of a discourse, the traditional distributional view is not sufficient. We argue that, to grow into an empirically adequate, full-fledged theory of meaning and interpretation, distributional models must evolve to provide meaning representations for actual language use in discourse and dialogue. Specifically, we discuss how the type of information they encode needs to be extended, and propose a series of tasks to evaluate pragmatically aware distributional models.

2   Meaning in Discourse

As we just pointed out, distributional semantics has been successful at providing data-driven meaning representations that are however limited to capturing generic, conceptual aspects of meaning. To use well established knowledge representation terms, distributional models capture the terminological knowledge (T-Box) of Description Logic, whereas they lack the encoding of assertional knowledge (A-Box), which refers to individuals (Brachman and Levesque, 1982). Proper natural language semantic modelling should capture both kinds of knowledge as well as their relation. Furthermore, distributional models have so far missed the main insights provided by the Dynamic Semantics tradition (Grosz et al., 1983;

Grosz and Sidner, 1986; Kamp and Reyle, 1993; Asher and Lascarides, 2003; Ginzburg, 2012), namely, that the meaning of an expression consists in its context-change potential, where context is incrementally built up as a discourse proceeds.

We contend that a distributional semantics for language use should account for the *discourse context-dependent*, *dynamic*, and *incremental* nature of language. Generic semantic knowledge won't suffice: one needs to encode somehow the discourse state or common ground, which will enable modeling discourse and dialogue coherence. In this section, we first look into examples that illustrate the dependence of interpretation on discourse and dialogue context and then consider the dynamic meaning of sentences as context-change potential.

2.1 Word and Phrase Meaning

As is well known, standard distributional models provide a single meaning representation for a word, which implicitly encodes all its possible senses and meaning nuances in general. A few recent models do account for some contextual effects within the scope of a sentence: For instance, the different shades of meaning that an adjective like *red* takes depending on the noun it modifies (e.g., *car* vs. *cheek*). However, such models, e.g. Erk and Padó (2008), Dinu and Lapata (2010), and Erk et al. (2013), typically use just a single word or sentence as context. They do not look into how word meaning gets progressively constrained by the common ground of the speakers as the discourse unfolds.

A prominent type of "meaning adjustment" in discourse and dialogue is the interaction with the properties of the referent a particular word is associated to. For example, when we use a word like *box*, which a priori can be used for entities with very different properties, we typically use it to refer to a specific box in a given context, and this constrains its interpretation. The referential effects extend to composition. Consider, for instance, the following example by McNally and Boleda (2015): Adrian and Barbara are sorting objects according to color in different, identical, brown cardboard boxes. Adrian accidentally puts a pair of red socks in the box containing blue objects, and Barbara remarks *'no, no, these belong in the red box'*. Thus, even if *red* when modifying *box* (or indeed any noun denoting a physical ob-

ject) will typically refer to its colour, it may also refer to other properties of the box referent (such as its contents) if these are prominent in the current discourse context and have become part of the common ground.

Indeed, the emergence of ad hoc meaning conventions in conversation is well attested empirically. In the classic psycholinguistic experiments by Clark and Wilkes-Gibbs (1986), speakers may first refer to a Tangram figure as *the one that looks like an angel* and end up using simply the word *angel* to mean that figure. As Garrod and Anderson (1987) point out, this idiosyncratic use of language "depends as much upon local and transient conventions, set up during the course of the dialogue, as [it does] on the more stable conventions of the larger linguistic community" (cf. Lewis (1969)). Arguably, current distributional models mainly capture the latter stable conventions. The challenge is thus to be able to also capture the former, discourse-dependent meaning.

Moreover, even function words, which are not-referential and are usually considered to have a precise (logical) meaning, are subject to pragmatic effects. For instance, the meaning of the determiner *some* is typically taken to be that of an existential quantifier (i.e., there exists at least one object with certain properties). Yet, its 'at least one' meaning may be refined in particular discourse contexts, as shown in the following examples:

(1) a. If you ate some of the cookies, then I won't have enough for the party.
    ↝ *some and possibly all*
  b. A: Did you eat all the cookies?
    B: I ate some. ↝ *some but not all*

Distributional models have so far not been particularly successful in modelling the meaning of function words (but see Baroni et al. (2012); Bernardi et al. (2013); Hermann et al. (2013)). We believe that discourse-aware distributional semantics may fare better in this respect. We elaborate on this idea further in the next subsection since their impact is seen beyond words and phrases level.

2.2 Beyond Words and Phrases

Following formal semantics, so far distributional semantics has modelled sentences as a product of a compositional function (Baroni et al., 2014; Socher et al., 2012; Paperno et al., 2014). The main focus has been on evaluating which compositional operation performs best against tasks

such as classifying sentence pairs in an entailment relation, evaluating sentence similarity (Marelli et al., 2014), or predicting the so-called "sentiment" (positive, negative, or neutral orientation) of phrases and sentences (Socher et al., 2013). None of these tasks have considered sentence pairs within a wider discourse or dialogue context.

We propose to take a different look on what the distributional meaning of a sentence is. Sentences are part of larger communicative situations and, as highlighted in the Dynamic Semantic tradition, can be considered relations between the discourse so far and what is to come next. We thus challenge the distributional semantics community to develop dynamic distributional semantic models that are able to encode the "context change potential" that sentences and utterances bring about as well as their coherence within a discourse context, including but not limited to anaphoric accessibility relations.

We believe that in this dynamic view function words will play a prominent role, since they have a large impact on how discourse unfolds. For instance, negation is known to generally block antecedent accessibility, as exemplified in (2a). Another example is presented in (2b) (see Paterson et al. (2011)): Speakers typically continue version (i) by mentioning properties of the reference set (e.g., *They listened carefully and took notes*), and (ii) by talking about the complement set (e.g., *They decided to stay at home instead*).

(2) a. It's not the case that John loves a woman$_i$. *She$_i$ is smart.
     b. (i) A few / (ii) Few of the students attended the lecture. They . . .

In the context of dialogue, adequate compositional distributional models should aim at capturing how an utterance influences the common ground of the dialogue participants (Stalnaker, 1978; Clark, 1996) and constrains possible follow-ups (Asher and Lascarides, 2003; Ginzburg, 2012). This requires taking into account the dialogue context, as exemplified in (3) from Schlöder and Fernández (2014), where the same utterance form (*Yes it is*) acts as either an acceptance (3a) or a rejection (3b) depending on its local context.

(3) a. A: But it's uh yeah it's an original idea.
         B: Yes it is.
     b. A: the shape of a banana is not- it's not really handy.
         B: Yes it is.

3 Tasks

Developing distributional semantic models that can tackle the phenomena discussed above is certainly challenging. However, we believe that, given the many recent advances in the field, the distributional semantics community is ready to take up this challenge. We have argued that, in order to account for the dynamics of situated common ground and coherence, it is critical to capture the discourse context-dependent and incremental nature of meaning. Here we sketch out a series of tasks related to some of the main phenomena we have discussed, against which new models could be evaluated.

In Section 2.1 we have considered the need to interface conceptual meaning with referential meaning incrementally built up as a discourse unfolds. A good testbed for evaluating these aspects is offered by the recent development of cross-modal distributional semantic frameworks that are able to map between language and vision (Karpathy et al., 2014; Lazaridou et al., 2014; Socher et al., 2014). Current models have shown that images representing a concept can be retrieved by mapping a word vector into a visual space, and more recently image generation systems that create images from word vectors have also been introduced (Lazaridou et al., 2015a; Lazaridou et al., 2015b). These frameworks could be used to test whether an incrementally constructed, discourse-contextualised word vector is able to retrieve and generate different, more contextually appropriate images than its out-of-context vector counterpart. For instance, a vector for a phrase like *red box* in a context where *red* refers to the box' contents should be mapped to different types of images depending on whether it has been constructed by a pragmatically aware model or not. Such a dataset could be constructed by creating images of referents of the same phrase used in different contexts, where the task would be to pick the best image for each context.

A related task would be reference resolution in a situated visual dialogue context (which can be seen as a situated version of image retrieval). This task has recently been tackled by Kennington and Schlangen (2015), who present an incremental ac-

count of word and phrase meaning with an approach outside the distributional semantics framework but very close in spirit to the issues we have discussed here. Given a representation of a referring expression and a set of visual candidate referents, the task consists in picking up the intended referent by incrementally processing and composing the words that make up the expression. Such a task (or versions thereof where contextual information beyond the referring expression is used) thus seems a good candidate for evaluating dynamic distributional models.

In Section 2.2, we have highlighted the context update potential of utterances as a feature that should be captured by compositional distributional models beyond the word/phrase level. Recent work has evaluated such models on dialogue act tagging tasks (Kalchbrenner and Blunsom, 2013; Milajevs et al., 2014). However, these approaches consider utterances in isolation and rely on a predefined set of dialogue act types that are to a large extent arbitrary, and in any case of a meta-linguistic nature. Similar comments would apply to the task of identifying discourse relations connecting isolated pairs of sentences. Instead, we argue that pragmatically-aware distributional models should help us to induce dialogue acts in an unsupervised way and to model them as context update functions. Thus, we suggest to adopt tasks that target coherence and the evolution of common ground — which is what discourse relations and dialogue acts are meant to convey in the first place — in a more direct way.

One possible task would be to assess whether (or the extent to which) an utterance is a coherent continuation of the preceding discourse. Another one would be to predict the next sentence or utterance. Simple versions of similar tasks have started to be addressed by recent approaches (Hu et al., 2014, among others), see Section 4 for discussion. We propose to adopt these tasks, namely *coherence ranking of possible next sentences* and *next sentence prediction*, to evaluate pragmatically aware compositional distributional semantic models. Given the crucial role that function words, as discussed above, play with respect to how the discourse can unfold, these tasks should include the effects of function words on discourse/dialogue continuation.

For the design of other concrete instances of these tasks, it would be worth to take into account the evaluation frameworks developed in the field of applied dialogue systems research (and thus outside the distributional semantics tradition) by Young et al. (2013), who have proposed probabilistic models that can compute distributions over dialogue contexts, and can thus to some extent predict (or choose) a next utterance.

4 Related Work

In this position paper we have focused on the shortcomings of existing standard distributional models regarding their ability to capture the dynamics of the discourse/dialogue context and their impact on meaning. Some models have aimed at capturing the word meaning of a specific word occurrence in context. These approaches offer a very valuable starting point, but their scope differs from ours. In particular, we can identify the following three main traditions: (1) Word Sense Disambiguation (Navigli, 2009, offers an overview), which aims to assign one of the predefined list of word senses to a given word, depending on the context. These are typically dictionary senses, and so do not capture semantic nuances that depend on the specific use of the word in a given discourse or dialogue context. (2) Word meaning in context as modeled in the lexical substitution task (McCarthy and Navigli, 2007; Erk et al., 2013), which predicts one or more paraphrases for a word in a given sentence. Unlike Word Sense Disambiguation, word meaning in context is specific to a given use of a word, that is, it doesn't assume a pre-defined list of senses and can account for highly specific contextual effects. However, in this tradition context is restricted to one sentence, so the semantic phenomena modeled do not extend to discourse or dialogue. (3) Compositional distributional semantics (Baroni and Zamparelli, 2010; Mitchell and Lapata, 2010; Boleda et al., 2013), which predicts the meaning of a phrase or sentence from the meaning of its component units. For instance, compositional distributional semantics accounts for how the generic distributional representation of, say, *red* makes different contributions when composed with nouns like *army*, *wine*, *cheek*, or *car*, by modeling the resulting phrase. However, these methods are again limited to intra-sentential context and only yield one single interpretation per phrase (presumably, the most typical one), thus not accounting for context-dependent interpretations of the *red box* type, discussed in

Section 2.1.

A few existing approaches can be seen as first steps towards a more discourse-aware distributional semantics, like the paper by McNally and Boleda (2015), which sketches a way to integrate compositional distributional semantics into Discourse Representation Theory (Kamp and Reyle, 1993). In addition, Herbelot (2015) has provided contextualized distributional representations for referential entities denoted by proper nouns in literary works. However, her procedure is still non-incremental in nature. Newer distributional models, such as Mikolov's SKIP-GRAM model (Mikolov et al., 2013), could incrementally update the representation of entities, and some work has been done in linking this model to the external world through images (Lazaridou et al., 2015c). However, these models do not yet account for specific, differentiated, discourse context-dependent interpretations of words of the sort discussed above, and they give a simple distributional representation of function words that does not readily account for their role in discourse.

Coherence ranking and sentence prediction, which we propose as the core testing ground, recently started being addressed, even if existing benchmarks have not been developed with the goals we highlighted above. The systems developed in Hu et al. (2014) have been successfully applied, among other things, to the task of choosing the correct response to a tweet, while Vinyals and Le (2015) and Sordoni et al. (2015) use neural models to *generate* responses for online dialogue systems and tweets, respectively (in the latter case taking into account a wider conversational context). These initial approaches are very promising, but they are disconnected from the referential context. Moreover, they have so far been trained specifically to achieve their goals, and it is not clear to what extent they can be integrated with a general semantic theory to serve other purposes.

Finally, the possibility of developing a pragmatically-oriented distributional semantics has been pointed out by Purver and Sadrzadeh (2015), who focus on opportunities for cross-fertilisation between dialogue research and distributional models. We certainly agree that the time is ripe for those and the other proposals made in this paper.

5 Conclusions

Distributional models are an important step towards building computational systems that can mimic human linguistic ability. However, we have argued that, as they stand, they still cannot account for language in use – that is, language within a discourse or a dialogue context, in a situated environment. We have described several linguistic phenomena that a comprehensive semantic model should account for, and proposed some concrete tasks that could serve to evaluate the adequacy of new-generation semantic systems targeting them. One crucial aspect that should be explored, however, is to what extent current distributional models need to be extended, and to what extent they need to be integrated into different frameworks, if the phenomena we have explored in this paper fall outside the distributional scope. We really hope that the community will take on this and the other challenges we have put forth in this paper.

Acknowledgements

References

Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.

Marco Baroni and Roberto Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of EMNLP*, pages 1183–1193, Boston, MA.

Marco Baroni, Raffaella Bernardi, Ngoc-Quynh Do, and Chung-chieh Shan. 2012. Entailment above the word level in distributional semantics. In *Proceedings of the 13th Conference of the European Chap-*

ter of the Association for Computational Linguistics, pages 23–32.

Marco Baroni, Raffaella Bernardi, and Roberto Zamparelli. 2014. Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technology*, 9(6):5–110.

Raffaella Bernardi, Georgiana Dinu, Marco Marelli, and Marco Baroni. 2013. A relatedness benchmark to test the role of determiners in compositional distributional semantics. In *Proceedings of ACL 2013 (Volume 2: Short Papers)*, pages 53–57.

Gemma Boleda, Marco Baroni, The Nghia Pham, and Louise McNally. 2013. Intensionality was only alleged: On adjective-noun composition in distributional semantics. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)*, pages 35–46.

Ronald J. Brachman and Hector J. Levesque. 1982. Competence in knowledge representation. In *Proceedings of AAAI*, pages 189–192.

Herbert H. Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22(1):1–39.

Herbert H. Clark. 1996. *Using language*. Cambridge University Press.

Georgiana Dinu and Mirella Lapata. 2010. Measuring distributional similarity in context. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1162–1172, Cambridge, MA, October. Association for Computational Linguistics.

Katrin Erk and Sebastian Padó. 2008. A structured vector space model for word meaning in context. In *Proceedings of EMNLP*, Honolulu, HI. To appear.

Katrin Erk, Diana McCarthy, and Nicholas Gaylord. 2013. Measuring Word Meaning in Context. *Computational Linguistics*, 39(3):511–554, September.

Katrin Erk. 2012. Vector space models of word meaning and phrase meaning: A survey. *Language and Linguistics Compass*, 6(10):635–653.

Simon Garrod and Anthony Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2):181–218.

Jonathan Ginzburg. 2012. *The Interactive Stance*. Oxford University Press.

Edward Grefenstette and Mehrnoosh Sadrzadeh. 2011. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of EMNLP*, pages 1394–1404.

Barbara J. Grosz and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational linguistics*, 12(3):175–204.

Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1983. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st annual meeting on Association for Computational Linguistics*, pages 44–50.

Aurélie Herbelot. 2015. Mr Darcy and Mr Toad, gentlemen: distributional names and their kinds. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 151–161. ACL.

Karl Moritz Hermann, Edward Grefenstette, and Phil Blunsom. 2013. "Not not bad" is not "bad": A distributional account of negation. In *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*, pages 74–82. ACL.

Baotian Hu, Zhengdong Lu, Hang Li, and Qingcai Chen. 2014. Convolutional neural network architectures for matching natural language sentences. In Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2042–2050. Curran Associates, Inc.

Nal Kalchbrenner and Phil Blunsom. 2013. Recurrent convolutional neural networks for discourse compositionality. In *Proceedings of the 2013 Workshop on Continuous Vector Space Models and their Compositionality*.

Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic*. Dordrecht: Kluwer.

Andrej Karpathy, Armand Joulin, and Li Fei-Fei. 2014. Deep Fragment Embeddings for Bidirectional Image Sentence Mapping. *Available at arXiv:1406.5679*.

Casey Kennington and David Schlangen. 2015. Simple learning and compositional application of perceptually grounded word meanings for incremental reference resolution. In *Proceedings of the Conference for the Association for Computational Linguistics (ACL)*.

Angeliki Lazaridou, Elia Bruni, and Marco Baroni. 2014. Is this a wampimuk? cross-modal mapping between distributional semantics and the visual world. In East Stroudsburg PA: ACL, editor, *Proceedings of ACL 2014*, pages 1403–1414.

Angeliki Lazaridou, Dat Tien Nguyen, Raffaella Bernardi, and Marco Baroni. 2015a. Unveiling the dreams of word embeddings: Towards language-driven image generation. Technical report, University of Trento. Submitted, available at arXiv:1506.03500.

Angeliki Lazaridou, Nghia Pham, and Marco Baroni. 2015b. Combining language and vision with a multimodal skip-gram model. In East Stroudsburg PA: ACL, editor, *Proceedings of NAACL HLT 2015*, pages 153–163.

Angeliki Lazaridou, Nghia The Pham, and Marco Baroni. 2015c. Combining language and vision with a multimodal skip-gram model. In *Proceedings of NAACL HLT 2015*, pages 153–163.

David Lewis. 1969. *Convention: A philosophical study*. Harvard University Press.

Marco Marelli, Luisa Bentivogli, Marco Baroni, Raffaella Bernardi, Stefano Menini, and Roberto Zamparelli. 2014. Semeval-2014 task 1: Evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 1–8.

Diana McCarthy and Roberto Navigli. 2007. Semeval-2007 task 10: English lexical substitution task. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 48–53, Prague, Czech Republic, June. Association for Computational Linguistics.

Louise McNally and Gemma Boleda. 2015. Conceptual vs. referential affordance in concept composition. In Y. Winter and J. Hampton, editors, *Concept Composition and Experimental Semantics/Pragmatics*. Springer. Accepted for publication.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.

Dmitrijs Milajevs, Dimitri Kartsaklis, Mehrnoosh Sadrzadeh, and Matthew Purver. 2014. Evaluating neural word representations in tensor-based compositional settings. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 708–719.

Jeff Mitchell and Mirella Lapata. 2010. Composition in distributional models of semantics. *Cognitive Science*, 34(8):1388–1429.

Roberto Navigli. 2009. Word sense disambiguation: A survey. *ACM Computing Surveys*.

Denis Paperno, Nghia The Pham, and Marco Baroni. 2014. A practical and linguistically-motivated approach to compositional distributional semantics. In *Proceedings of ACL*, pages 90–99, Baltimore, MD.

Kavin Paterson, Ruth Filik, and Linda Moxey. 2011. Quantifiers and discourse processing. *Language and Linguistics Compass*, pages 1–29.

Matthew Purver and Mehrnoosh Sadrzadeh. 2015. From distributional semantics to distributional pragmatics? In *Proceedings of the IWCS 2015 Workshop on Interactive Meaning Construction*, pages 21–22, London, UK.

Julian J. Schlöder and Raquel Fernández. 2014. The role of polarity in inferring acceptance and rejection in dialogue. In *Proceedings of the SIGdial 2014 Conference*.

Richard Socher, Brody Huval, Christopher D Manning, and Andrew Y Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of EMNLP-CNLL*, pages 1201–1211.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642.

Richard Socher, Andrej Karpathy, Quoc V. Le, Christopher D. Manning, and Andrew Y. Ng. 2014. Grounded Compositional Semantics for Finding and Describing Images with Sentences. *Transactions of the Association for Computational Linguistics*.

Alessandro Sordoni, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Meg Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan. 2015. A neural network approach to context-sensitive generation of conversational responses. Conference of the North American Chapter of the Association for Computational Linguistics Human Language Technologies (NAACL-HLT 2015), June.

Robert Stalnaker. 1978. Assertion. In P. Cole, editor, *Pragmatics*, volume 9 of *Syntax and Semantics*, pages 315–332. New York Academic Press.

Peter D Turney, Patrick Pantel, et al. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37(1):141–188.

Oriol Vinyals and Quoc V. Le. 2015. A neural conversational model. *CoRR*, abs/1506.05869.

Steve Young, Milica Gasic, Blaise Thomson, and Jason Williams. 2013. Pomdp-based statistical spoken dialogue systems: a review. *Proceedings of the IEEE*, PP(99):1–20.