How to make ends meet: Why general purpose NLU needs linguistics

Emily M. Bender
University of Washington

Relevance of Linguistic Structure in Neural NLP
ACL 2018 Workshop
19 July 2018

Claims

1. Human language is a general purpose communication tool

2. Work on NLP (neural and otherwise) isn't yet. Instead, two types:

    A. Task-focused and task-specific

    B. Linguistic-structure focused

3. General purpose NLU will require learning or building a system that can map between strings and linguistic meanings

--- i.e. something like a grammar

Sentence meaning, speaker meaning

- Learning correlations between domain-typical surface forms and task-specific representations conflates:
  - timeless/conventional/standing/sentence meaning
  - utterer/occasion/speaker meaning
- Drawbacks:
  - resolving the same problems around grammatical structure for each task
  - unlikely to scale to general-purpose NLU

Meaning derived from form is different from meaning derived from context of use

- Meaning level 1: Semantics derived from form: What is constant about the meaning of a sentence across different occasions of use

A: Is it raining?
B: Yes.

Meaning derived from form is different from meaning derived from context of use

- Meaning level 2: What a speaker publicly commits to by virtue of using a particular form in a particular context

A: Is it raining?
B: Yes.
A: It's perfectly dry outside. You're lying.
B: #I didn't say it's raining.

Meaning derived from form is different from meaning derived from context of use

- Meaning level 3: Inferences about a speaker's private cognitive states, but which the speaker hasn't publicly committed to

A: Is it raining?
B: Yes.
A: Oh, so you do think I should take my umbrella.
B: I didn't say that.

(Lascarides & Asher 2009, Asher & Lascarides 2013)

Leveraging sentence meaning

- Outside of artificial annotated data sets, machines don't have access to any direct representation of speaker meaning, only to natural language utterances
  - And the artificial, annotated data sets include only specific subsets of meaning
- Sentence meaning doesn't determine situated speaker meaning, but is an important cue to it (Quine 1960, Grice 1968, Reddy 1979, Clark 1996)
- A task-independent, comprehensive representation of sentence meaning capturing exactly the information in the linguistic signal itself should benefit NLU systems
- … and is critical for general-purpose NLU

My goals

- Make NLP researchers aware of the sentence meaning/speaker meaning distinction and importance of sentence meaning

- Find ways to effectively bring linguistic knowledge to both NLP systems and NLP research

NAACL 2018 snapshot

- 50 NAACL 2018 long papers (#1001-1050)
- Surveyed for:
  - Task
  - NN input & output
  - Representation of natural language syntax
  - Notion of meaning
  - Language studied

All neural nets?

- 45/50 papers were primarily neural-based methods
- Two not at all: Amorim et al 2018, Cocos et al 2018
- Three minimally (included pretrained embeddings only): Gupta et al 2018, Ziai & Meurers 2018, Kriz et al 2018

Languages: Language pairs

- Ainu > English
- English > French
- English > German
- English > Vietnamese
- German > English
- Hungarian > English
- Korean > English
- Mboshi > French
- Spanish > English

Languages

- 12 tasks: English
- 5 tasks: German
- 3 tasks: Italian
- 2 tasks: Japanese, Chinese, Portuguese, Russian, Turkish, Czech
- 1 task: Finnish, French, Greek, Hausa, Latvian, Mexicanero, Nahuatl, Romanian, Spanish, Tamil, Urdu, Wixarika, Yorem Nokki
- n/a: 1 task
- massively multilingual: 1 task

42 tasks: Unnamed mystery language

Tasks: Concerned with form

- *Generative model of vowel typology* (Cotterell & Eisner 2018)
- *Morphological segmentation* (Kann et al 2018)
- *Grammatical error detection* (Rei & Søgaard 2018)

Tasks: Text transformation

- *MT* (Passban et al 2018, Anastasopoulos & Chiang 2018, Nguyen & Chiang 2018, Gu et al 2018, Edunov et al 2018)
- *Paraphrase detection/identification* (Kiela et al 2018, Issa et al 2018, Pagliardini et al 2018)
- *Text simplification* (Ma et al 2018, Kriz et al 2018, Vulić et al 2018)
- *Summarization* (Ma et al 2018, Edunov et al 2018)
- *Style transfer* (Rao & Tetreault 2018)
- *Rephrasing based on marked portions* (Grangier & Auli 2018)

Tasks: Generation

- *Coherent text generation* (Bosselut et al 2018)
- *Lyric generation fit to melody* (Watanabe et al 2018)
- *NLG from dialogue act representation* (Juraska et al 2018)
- *Question generation* (Elsahar et al 2018)

Tasks: Analyze linguistic structure (not task specific)

- *Focus/background labeling* (Ziai & Meurers 2018)
- *Parsing conversational speech* (Tran et al 2018)
- *Discourse relation labeling (explicit & implicit)* (Dai & Huang 2018)
- *Possession relation detection & classification* (Chinnappa & Blanco 2018)
- *Dialogue state tracking* (Vulić et al 2018)
- *Hedge detection* (Rei & Søgaard 2018)
- *Medical NER* (Wang et al 2018b)
- *Fine-grained entity type classification* (Xu & Barbosa 2018)

Tasks: Word-sense similarity

- *Synonym detection* (Jana & Goyal 2018)
- *Word relatedness* (Jana & Goyal 2018)
- *Word similarity* (Petroni et al 2018, Jana & Goyal 2018, Vulić et al 2018)
- *Analogy completion* (Jana & Goyal 2018)
- *Detecting word sense change* (Rosenfeld & Erk 2018)

Tasks: Text classification

- *Citation suggestion* (Bhagavatula et al 2018)
- *Microblog recommendation* (Zeng et al 2018)
- *Automated essay scoring* (Amorin et al 2018)
- *Automated essay scoring with adversarial input* (Farag et al 2018)
- *Text classification* (KM et al 2019, Petroni et al 2018)
- *Topic modeling* (Benton & Dredze 2018)

Tasks: Sentiment analysis

- *Multi-domain sentiment analysis* (Liu et al 2018)
- *Sentiment (varied)* (Kiela et al 2018, Pagliardini et al 2018)
- *Token-level sentiment* (Rei & Søgaard 2018)

Tasks: Identifying task-specific speaker meaning

- *Parsing to scene graphs* (Wang et al 2018a)
- *Knowledge graph completion* (Ishihara et al 2018)
- *Variable typing* (Stathopoulos et al 2018)
- *Predicting which comments will change someone's mind* (Jo et al 2018)
- *Complex cross-session task extraction* (Sen et al 2018)
- *Ad hominem attack recognition* (Habernal et al 2018)
- *Predicting applause in political speeches* (Gillick et al 2018)

Tasks: QA/IE/IR

- *Better datasets for Visual QA* (Chao et al 2018)
- *Quantifier selection given scene* (Pezzelle et al 2018)
- *QA w/heterogenous memory* (Fu & Feng 2018)
- *Question type classification* (Pagliardini et al 2018)
- *Relation extraction* (Gupta et al 2018)
- *Taxonomy construction* (Cocos et al 2018)
- *Unsupervised hypernym detection* (Chang et al 2018)

Tasks: Inference/entailment

- *STS, SNLI, SICK* (KM et al 2018, Kiela et al 2018, Pagliardini et al 2018)
- *Multi sentence reading comprehension* (Khashabi et al 2018)
- *Logical reasoning (not actually language though)* (Ishihara et al 2018)

Representations of NL syntax

- Templates (1)
- Unknown (2) or n/a (4)
- POS tag embeddings (1)
- constituent structure (2)
- dependencies (5)
- constituent structure and dependencies (1)
- none (48)

Representation of NL semantics

- Embeddings: word, n-gram, sentence, affix, stem, character, doc, entity, clique
- Enhancements to embeddings: grounding, memory network, time-inclusive, cross-linguistic
- Other: non-neural vector-space representations, AMR, RST labels, WordNet, PPDB, slot-value pairs

Notions of meaning

- Images
- Paraphrase relations
- Translational equivalence
- Sentiment
- Word-sense similarity
- Information structure
- Knowledge graphs
- Scene graphs
- Search intent
- Argumentation strategy
- Rhetorical tension & closure
- Truth of an answer given an image
- Inference
- Type of variables in equations
- Perlocutionary acts: Changes in view

Sentence meaning, speaker meaning

- Learning correlations between domain-typical surface forms and task-specific representations conflates:
    - timeless/conventional/standing/sentence meaning
    - utterer/occasion/speaker meaning
- Drawbacks:
    - resolving the same problems around grammatical structure for each task
    - unlikely to scale to general-purpose NLU

Why bother with sentence meaning?

- Claim: Sentence meaning, but not speaker meaning, is compositional
- Claim: Systems attempting to understand speaker meaning would benefit from reusable, automatically derivable, task-independent representations of sentence meaning
- Furthermore: A compositional approach to creating sentence meaning representations provides
  - Comprehensiveness
  - Consistency
  - Scalability

(Bender et al 2015)

A meaning representation system is compositional if (working definition):

- it is grounded in a finite (possibly large) number of atomic symbol-meaning pairings
- it is possible to create larger symbol-meaning pairings by combining the atomic pairings through a finite set of rules;
- the meaning of any non-atomic symbol-meaning pairing is a function of its parts and the way they are combined;
- this function is possibly complex, containing special cases for special types of syntactic combination, but only draws on the immediate constituents and any semantic contribution of the rule combining them; and
- further processing will not need to destructively change a meaning representation created in this way to create another of the same type.

(Bender et al 2015)

Semantic annotation survey: Compositional layer

- Predicate-argument structure
- Partial constraints on:
  - Scope of negation and other operators
  - Restriction of quantifiers
  - Modality
  - Tense/aspect/mood
  - Information structure
  - Discourse status of referents of NPs
  - Politeness
- Possibly compositional, but not according to sentence grammar:
  - Coherence relations/rhetorical structure

(Bender et al 2015)

Importance of syntax in recovering sentence meaning

先生によると男の子よりも女の子がポケモンがすきだ。[jpn]

先生 によると 男の子 よりも 女の子 が ポケモン が すき だ。

teacher (.) boy (.) girl (.) Pokemon (.) like (.)

teacher ACCORDING.TO boy THAN girl NOM Pokemon NOM like COP.PRES

Claims

1. Human language is a general purpose communication tool

2. Work on NLP (neural and otherwise) isn't yet. Instead, two types:

    A. Task-focused and task-specific

    B. Linguistic-structure focused

3. General purpose NLU will require learning or building a system that can map between strings and linguistic meanings

--- i.e. something like a grammar

My goals

- Make NLP researchers aware of the sentence meaning/speaker meaning distinction and importance of sentence meaning

- Find ways to effectively bring linguistic knowledge to both NLP systems and NLP research

Bringing linguistics to NLP research

- Accessible textbooks (*Linguistic Fundamentals for Natural Language Processing*, Bender 2013, Bender & Lascarides forthcoming)
- Reviewing practices (COLING 2018)
- Error analysis best practices
- Build It Break It shared tasks
- Large-scale encodings of rich linguistic knowledge
  - Case to highlight: English Resource Grammar (Flickinger 2000) and its associated treebanks/sembanks (Oepen et al 2004, Flickinger et al 2017)

http://moin.delph-in.net/ErsTutorial

Is there even any structure in natural language?

- An actual question asked of me by a DL researcher working on an conversation agent

- Was attacking seq2seq problems and reasoning from neural architectures to hypotheses about what babies do

- Apparently convincing example: *Kim danced and Sandy sang in the park.*

Making ends meet

- A pile of end-to-end tasks does not a general purpose system make
- Does true machine NLU require a hand-built grammar?
  - Not necessarily
  - But my bet is some kind of grammar (built or learned) will be required
  - And the way we're currently setting up the NLU-related machine learning tasks isn't generally leading to grammar learning
  - NLP needs linguists, to better understand the shape of the problem and to better evaluate progress

References

Amorim, E., Cançado, M., and Veloso, A. (2018). Automated essay scoring in the presence of biased ratings. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 229–237. Association for Computational Linguistics.

Anastasopoulos, A. and Chiang, D. (2018). Tied multitask learning for neural speech translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 82–91. Association for Computational Linguistics.

Asher, N. and Lascarides, A. (2013). Strategic conversation. *Semantics and Pragmatics*, **6**(2), 2:1–:62.

Bender, E. M., Flickinger, D., Oepen, S., Packard, W., and Copestake, A. (2015). Layers of interpretation: On grammar and compositionality. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 239–249, London, UK. Association for Computational Linguistics.

Benton, A. and Dredze, M. (2018). Deep dirichlet multinomial regression. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 365–374. Association for Computational Linguistics.

Bhagavatula, C., Feldman, S., Power, R., and Ammar, W. (2018). Content-based citation recommendation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 238–251. Association for Computational Linguistics.

Bosselut, A., Celikyilmaz, A., He, X., Gao, J., Huang, P.-S., and Choi, Y. (2018). Discourse-aware neural rewards for coherent text generation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 173–184. Association for Computational Linguistics.

Chang, H.-S., Wang, Z., Vilnis, L., and McCallum, A. (2018). Distributional inclusion vector embedding for unsupervised hypernymy detection. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 485–495. Association for Computational Linguistics.

Chao, W.-L., Hu, H., and Sha, F. (2018). Being negative but constructively: Lessons learnt from creating better visual question answering datasets. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 431–441. Association for Computational Linguistics.

Chinnappa, D. and Blanco, E. (2018). Mining possessions: Existence, type and temporal anchors. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 496–505. Association for Computational Linguistics.

Cocos, A., Apidianaki, M., and Callison-Burch, C. (2018). Comparing constraints for taxonomic organization. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 323–333. Association for Computational Linguistics.

Cotterell, R. and Eisner, J. (2018). A deep generative model of vowel formant typology. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 37–46. Association for Computational Linguistics.

Dai, Z. and Huang, R. (2018). Improving implicit discourse relation classification by modeling interdependencies of discourse units in a paragraph. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 141–151. Association for Computational Linguistics.

Edunov, S., Ott, M., Auli, M., Grangier, D., and Ranzato, M. (2018). Classical structured prediction losses for sequence to sequence learning. In *Proceedings of the 2018 Conference of the North American*

Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pages 355–364. Association for Computational Linguistics.

Elsahar, H., Gravier, C., and Laforest, F. (2018). Zero-shot question generation from knowledge graphs for unseen predicates and entity types. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 218–228. Association for Computational Linguistics.

Farag, Y., Yannakoudakis, H., and Briscoe, T. (2018). Neural automated essay scoring and coherence modeling for adversarially crafted input. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 263–271. Association for Computational Linguistics.

Flickinger, D. (2000). On building a more efficient grammar by exploiting types. *Natural Language Engineering*, **6 (1) (Special Issue on Efficient Processing with HPSG)**, 15 – 28.

Flickinger, D., Oepen, S., and Bender, E. M. (2017). Sustainable development and refinement of complex linguistic annotations at scale. In N. Ide and J. Pustejovsky, editors, *Handbook of Linguistic Annotation*, pages 353–377. Springer Netherlands, Dordrecht.

Fu, Y. and Feng, Y. (2018). Natural answer generation with heterogeneous memory. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 185–195. Association for Computational Linguistics.

Gillick, J. and Bamman, D. (2018). Please clap: Modeling applause in campaign speeches. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 92–102. Association for Computational Linguistics.

Grangier, D. and Auli, M. (2018). Quickedit: Editing text & translations by crossing words out. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 272–282. Association for Computational Linguistics.

Gu, J., Hassan, H., Devlin, J., and Li, V. O. (2018). Universal neural machine translation for extremely low resource languages. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 344–354. Association for Computational Linguistics.

Gupta, P., Roth, B., and Schütze, H. (2018). Joint bootstrapping machines for high confidence relation extraction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 26–36. Association for Computational Linguistics.

Habernal, I., Wachsmuth, H., Gurevych, I., and Stein, B. (2018). Before name-calling: Dynamics and triggers of ad hominem fallacies in web argumentation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 386–396. Association for Computational Linguistics.

Ishihara, T., Hayashi, K., Manabe, H., Shimbo, M., and Nagata, M. (2018). Neural tensor networks with diagonal slice matrices. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 506–515. Association for Computational Linguistics.

Issa, F., Damonte, M., Cohen, S. B., Yan, X., and Chang, Y. (2018). Abstract meaning representation for paraphrase detection. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 442–452. Association for Computational Linguistics.

Jana, A. and Goyal, P. (2018). Can network embedding of distributional thesaurus be combined with word vectors for better representation? In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 463–473. Association for Computational Linguistics.

Jo, Y., Poddar, S., Jeon, B., Shen, Q., Rose, C., and Neubig, G. (2018). Attentive interaction model: Modeling changes in view in argumentation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 103–116. Association for Computational Linguistics.

Juraska, J., Karagiannis, P., Bowden, K., and Walker, M. (2018). A deep ensemble model with slot alignment for sequence-to-sequence natural language generation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 152–162. Association for Computational Linguistics.

K M, A., Basu Roy Chowdhury, S., and Dukkipati, A. (2018). Learning beyond datasets: Knowledge graph augmented neural networks for natural language processing. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 313–322. Association for Computational Linguistics.

Kann, K., Mager Hois, J. M., Meza Ruiz, I. V., and Schütze, H. (2018). Fortification of neural morphological segmentation models for polysynthetic minimal-resource languages. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 47–57. Association for Computational Linguistics.

Khashabi, D., Chaturvedi, S., Roth, M., Upadhyay, S., and Roth, D. (2018). Looking beyond the surface: A challenge set for reading comprehension over multiple sentences. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 252–262. Association for Computational Linguistics.

Kiela, D., Conneau, A., Jabri, A., and Nickel, M. (2018). Learning visually grounded sentence representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 408–418. Association for Computational Linguistics.

Kriz, R., Miltsakaki, E., Apidianaki, M., and Callison-Burch, C. (2018). Simplification using paraphrases and context-based lexical substitution. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 207–217. Association for Computational Linguistics.

Lascarides, A. and Asher, N. (2009). Agreement, disputes and commitment in dialogue. *Journal of Semantics*, **26**(2), 109–158.

Liu, Q., Zhang, Y., and Liu, J. (2018). Learning domain representation for multi-domain sentiment classification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 541–550. Association for Computational Linguistics.

Ma, S., Sun, X., Li, W., Li, S., Li, W., and Ren, X. (2018). Query and output: Generating words by querying distributed word representations for paraphrase generation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 196–206. Association for Computational Linguistics.

Nguyen, T. and Chiang, D. (2018). Improving lexical choice in neural machine translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 334–343. Association for Computational Linguistics.

Oepen, S., Toutanova, K., Shieber, S., Manning, C., Flickinger, D., and Brants, T. (2002). The LinGO Redwoods treebank. Motivation and preliminary applications. In *Proceedings of the 19th International Conference on Computational Linguistics*, Taipei, Taiwan.

Pagliardini, M., Gupta, P., and Jaggi, M. (2018). Unsupervised learning of sentence embeddings using compositional n-gram features. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 528–540. Association for Computational Linguistics.

Passban, P., Liu, Q., and Way, A. (2018). Improving character-based decoding using target-side morphological information for neural machine translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 58–68. Association for Computational Linguistics.

Petroni, F., Plachouras, V., Nugent, T., and Leidner, J. L. (2018). attr2vec: Jointly learning word and contextual attribute embeddings with factorization machines. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 453–462. Association for Computational Linguistics.

Pezzelle, S., Sorodoc, I.-T., and Bernardi, R. (2018). Comparatives, quantifiers, proportions: a multi-task model for the learning of quantities from vision. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 419–430. Association for Computational Linguistics.

Rao, S. and Tetreault, J. (2018). Dear sir or madam, may i introduce the gyafc dataset: Corpus, benchmarks and metrics for formality style transfer. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 129–140. Association for Computational Linguistics.

Rei, M. and Søgaard, A. (2018). Zero-shot sequence labeling: Transferring knowledge from sentences to tokens. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 293–302. Association for Computational Linguistics.

Rosenfeld, A. and Erk, K. (2018). Deep neural models of semantic shift. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 474–484. Association for Computational Linguistics.

Sen, P., Ganguly, D., and Jones, G. (2018). Tempo-lexical context driven word embedding for cross-session search task extraction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 283–292. Association for Computational Linguistics.

Stathopoulos, Y., Baker, S., Rei, M., and Teufel, S. (2018). Variable typing: Assigning meaning to variables in mathematical text. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 303–312. Association for Computational Linguistics.

Tran, T., Toshniwal, S., Bansal, M., Gimpel, K., Livescu, K., and Ostendorf, M. (2018). Parsing speech: a neural approach to integrating lexical and acoustic-prosodic information. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 69–81. Association for Computational Linguistics.

Vulić, I., Glavaš, G., Mrkšić, N., and Korhonen, A. (2018). Post-specialisation: Retrofitting vectors of words unseen in lexical resources. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 516–527. Association for Computational Linguistics.

Wang, Y.-S., Liu, C., Zeng, X., and Yuille, A. (2018a). Scene graph parsing as dependency parsing. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 397–407. Association for Computational Linguistics.

Wang, Z., Qu, Y., Chen, L., Shen, J., Zhang, W., Zhang, S., Gao, Y., Gu, G., Chen, K., and Yu, Y. (2018b). Label-aware double transfer learning for cross-specialty medical named entity recognition. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1–15. Association for Computational Linguistics.

Watanabe, K., Matsubayashi, Y., Fukayama, S., Goto, M., Inui, K., and Nakano, T. (2018). A melody-conditioned lyrics language model. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1*

(Long Papers), pages 163–172. Association for Computational Linguistics.

Xu, P. and Barbosa, D. (2018). Neural fine-grained entity type classification with hierarchy-aware loss. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 16–25. Association for Computational Linguistics.

Zeng, X., Li, J., Wang, L., Beauchamp, N., Shugars, S., and Wong, K.-F. (2018). Microblog conversation recommendation via joint modeling of topics and discourse. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 375–385. Association for Computational Linguistics.

Ziai, R. and Meurers, D. (2018). Automatic focus annotation: Bringing formal pragmatics alive in analyzing the information structure of authentic data. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 117–128. Association for Computational Linguistics.