Neural Networks and Textual Inference: How did we get here and where do we go now?

Ignacio Cases and Lauri Karttunen

September 8, 2017

1 Introduction

An essential part of our understanding of natural language is the ability to infer from a statement or a larger piece of text what other things a reader would be entitled or inclined to conclude about what the author believes in addition to what she actually says. The field of NLP includes decades of work dedicated to this issue. In this section we discuss briefly the four mile posts that have led us to the point at which we are today and what we plan to do next:

1. the FraCaS data set from 1996
2. the series of RTE challenges from 2005 up to 2011
3. the recent SICK, SNLI and MultiNLI corpora
4. our own work on implicative constructions
5. the goals for this work group

1.1 FraCaS

The beginnings of the modern approach to the problem of textual inference can be traced to the FraCaS project (Cooper et al., 1996), a consortium sponsored by the European Union. The main contribution of the project was a set of 346 inference problems. The FraCaS problem set was strongly influenced by the fact that the senior members on the team (Robin Cooper, Hans Kamp, Manfred Pinkal, Stephen Pulman, and Jan van Eijck) were semanticists steeped in the Montague tradition. Consequently, the problem set has examples of a wide variety of problems in formal semantics: *generalized quantifiers*, *negation*, *monotonicity*, *anaphora*, *ellipsis*, *comparatives*, *adjectives*, *temporal relations*, and *propositional attitudes*. The problem set consists examples with one or two premises, a question, a hypothesis and a judgement about the relation between the premise(s) and the hypothesis. For example,

`fracas-006 answer: no`
P1: No really great tenors are modest.
Q: Are there really great tenors who are modest?
H: There are really great tenors who are modest.

`fracas-010 answer: yes`
P1: Most great tenors are Italian.
Q: Are there great tenors who are Italian?
H: There are great tenors who are Italian.

`fracas-012 answer: undef`[1]
P1: Few great tenors are poor.
Q: Are there great tenors who are poor?
H: There are great tenors who are poor.
A: Not many

One lasting contribution of FraCaS was to frame natural language inference as a three-way classification task: *yes* (= entails), *no* (= contradicts), and = *undefined* (neither, permits).

In the early versions of RTE (Dagan et al., 2006), the next attack of the problem, there was a regression into a two-way distinction: TRUE (= entails) and FALSE (= doesn't entail) but in later versions of the RTE test sets the three-way distinction of *entails*, *contradicts*, or *neither* was reintroduced thanks to arguments by FraCaS alumni such as Dick Crouch at PARC who pointed out that it was important to not to conflate the contradiction of A and B with the case of A and B entailing nothing at all of the truth or falsity of the other.

Another, a more controversial aspect of FraCaS, is that it that the semantic relations it postulates between premises and hypotheses are only based on the semantics of the particular construction and the lexical meaning of the words involved. The data set contains no examples where the answers would depend on augmenting the premise with some background knowledge about the world. For example, it was assumed that the antecedent of an anaphoric expression such as *he*, *it*, and *she* was included in the premise, the pronoun was supposed not to be deontic or refer to something salient in the context that was not explicitly mentioned.

The FraCaS data set was intended to be a testsuite for semantics, analogous to syntactic testsets that were created for evaluating the coverage of grammar implementations of theories such as HPSG and LFG. However, the project did not include any evaluations

[1]It is true that *few tenors are poor* can be considered true in the case where there are no poor tenors. However, unless the author qualifies the premise by saying *few tenors, in fact, no tenors,are poor*, the addressee would be justified in inferring that the author thinks that some tenors are poor. If the author were in the position to make a stronger statement, she should have said *no* instead of *few*. We will come back to this topic in the discussion of invited inferences.

and there was no follow-up. The first study that used a part of the FraCaS suite for its intended purpose appears to be MacCartney and Manning (2007).

1.2 RTE – Recognizing Textual Entailment Challenges

The influential initiative of Dagan et al. (2006) spawned a series of RTE workshops in which several research groups where competing on a shared task. For each RTE workshop, two data sets were prepared, one for training and development, the other for testing. As in the FraCaS suite, each example was a triple consisting of a premise T, one or more sentences, a hypothesis H, and a label L for the relation between T and H. The first RTE challenges were a two-way classification tasks. The correct label was either TRUE or FALSE depending on whether H was 'entailed' by T. For Dagan et al. (2006) this notion is not a logical entailment but a less strict relation:

> We say that T entails H if the meaning of H can be inferred from the meaning of T, as would typically be interpreted by people. This somewhat informal definition is based on (and assumes) common human understanding of language as well as common background knowledge.

One example of a plausible inference that is not a logical entailment is the triple

> `id="586" value="TRUE" task="QA"`
> T: The two suspects belong to the 30th Street gang, which became embroiled in one of the most notorious recent crimes in Mexico: a shootout at the Guadalajara airport in May, 1993, that killed Cardinal Juan Jesus Posadas Ocampo and six others.
> H: Cardinal Juan Jesus Posadas Ocampo died in 1993.[2]

Currently many people prefer the more general term *Natural Language Inference*, NLI, over *Recognizing Textual Entailment*, RTE.

The nature of the RTE data sets is the opposite of carefully constructed, small examples of FraCaS that typically illustrate the semantics of one particular construction or lexical item, such as the determiners *no*, *most*, and *few* in the examples above. The examples contain snippets of newspaper text, headlines. They are often challenging to parse and predicting the correct label may require in principle an unlimited amount of knowledge about the world. In RTE1 they range from trivial examples like

> `id="78" value="FALSE" task="IR"`[3]
> T: Clinton's new book is not big seller here.
> H: Clinton's book is a big seller.

[2] The cardinal might not have succumbed to his injuries right away.

[3] Logically this example is not a contradiction although some people might draw that conclusion. In a data set for three-way classification (*entails*, *contradicts*, *neither*), this example probably should be labeled *neither*. Deleting *here* in a downward monotonic environment is not a truth-preserving operation. The book could be a big seller everywhere else.

to truly difficult ones such as

> id="85" value="TRUE" task="IR"
> T: The country's largest private employer, Wal-Mart Stores Inc., is being sued by a number of its female employees who claim they were kept out of jobs in management because they are women.
> H: Wal-Mart sued for sexual discrimination.

This example requires understanding that *keep someone out of a job* is here a paraphrase of *not hire someone for a job*, and the knowledge that **not hiring someone for a job because of the applicant's gender** is an act of *sexual discrimination*, which is against the law.

Because of many difficult examples like this one, it is not surprising the the results of the RTE1 challenge are rather modest. Of the systems that covered all of the test data, the winner achieved 0.6 accuracy, only 0.1 above the baseline of randomly choosing between TRUE and FALSE. This left much room for further improvements.

The later rounds of RTE challenges deliver better results by using more of available linguistic resources, sometimes in a hacky way. The best result ever in the classical RTE task seems to have been achieved by Hickl and Bensley's (2007) system with its 0.8 accuracy in the binary version of the classification task. The system has a complex architecture that incorporates a multitude of components including *WordNet*, *n-gram\word similarity*, *Logical inference*, *ML Classification*, *Anaphora resolution*, *Entailment Corpora*, DIRT, and *Background Knowledge* (Giampiccolo et al., 2007). It is difficult to figure out from the architecture diagram in Figure 1 how the various

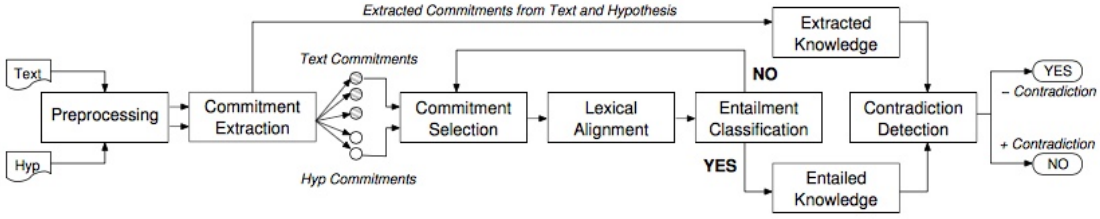

Figure 1: Hickl-Bensley System Architecture

lingware components interact with each other. But the overall strategy is clear: collect all the specific commitments from the text and the hypothesis separately, compare the list of T and H commitments at the end, and predict *contradicts* or *entails* accordingly.[4]

The data set in RTE 3 consisted of 1600 T-H pairs. Although it is nearly five times larger than the FRACAS suite, the systems derived from it or any other of the RTE challenge data sets does not have sufficient coverage for real world applications.

[4]I feel empathy for this approach because it is rather similar in overall design to the Bridge system (Bobrow et al., 2007) at PARC that were working at about the same time except that we started with a preprocessing step that converted the input into a knowledge representation. At the end of the Bridge pipe line we also had a component corresponding to the *Contradiction Detection* box in Figure 1.

Contradiction detection is a very difficult task. Consider the colors *blue* and *red*. A car that is blue is not red: *my car is blue* and *my car is red* is a contradiction. However, *I have a blue car* does not contradict *I have a red car*. But the statements *A blue car won the Indianapolis 500* and *A red car won the Indianapolis 500* cannot both be true about the same race. Replacing *won* by *lost* removes the contradiction. A winner is unique, losers are not.

If the phenomena to be covered is extended, a system such as Hickl and Bensley's cannot be scaled up manually. There is a huge number of special expressions similar to *sexual discrimination* that are difficult to define explicitly by logical inference rules although they might might be learnable sufficiently well given a large set of examples. The meaning of such multiword expressions cannot be defined compositionally or by a simple computable rule. Not finding any contradiction, Hickl and Bensley probably would correctly predict *entails* for the example above but it would incorrectly do the same for *Wal-Mart sued for sexual harassment*.

1.3 The recent SICK, SNLI, and MultiNLI corpora

Vector-based models for compositional semantics had been proposed in several articles (Mitchell and Lapata, 2008; Baroni and Zamparelli, 2010; Grefenstette and Sadrzadeh, 2011) but the progress was hampered by the lack of an RTE like corpus for training and evaluating such models. The SICK corpus (Sentences Involving Compositional Knowledge) was created for a shared task in SemEval-2014 (Marelli et al., 2014). It has about 10,000 T- H examples annotated for similarity and the semantic relation (*entailment*, *contradiction*, *neither*). The SICK examples were derived from descriptions of images and video snippets created by Turkers. From each seed sentence up to three new sentences were created manually: (i) a sentence with a similar meaning, (ii) a sentence with a contradictory meaning, and (iii) a sentence with mostly the same lexical items but a different meaning. In this step, 'syntactic and lexical transformations with predictable effects were applied' (Marelli et al. 2014: 2).

The purpose of SICK was to create a compositional semantics suite that did not require named-entity recognition or encyclopedic knowledge about the world. The semantic relation between T and H was meant to be decidable on purely linguistic evidence. In this respect the SICK data set is easier than the RTE test suites. The accuracy of the five best performing systems in SemEval-2014 were all above 80%.

Instead of concentrating on the application compositional techniques to word vectors, the SemEval-14 participants for the most part used non-compositional methods and resources such as WordNet. The best-performing systems combine multiple similarity metrics to infer the relation between the premise text and the hypothesis.

By the following year, the focus of the field had moved on to newer methods. The SNLI corpus (Bowman et al., 2015) was created to overcome the perception that the SICK data set was too small for developing neural network based models. SNLI is about 57 times larger than SICK, both are based on image captions. The first neural net model (Bowman et al., 2015) achieved 77.6% accuracy on the SNLI test set. In later models e.g. (Rocktäschel et al., 2015; Chen et al., 2017) the accuracy exceeds 80%.

The seed examples for SICK and SNLI were captions provided by Turkers for images. In developing the SICK corpus, the extension steps, (i)-(iii) above, for the seed image captions were created by linguists, by the developers of the data set. In the case of SNLI, the extension steps were outsourced to Turkers. Deriving the sentences from image captions has an unfortunate consequence: in these data sets contradiction is not semantic contradiction. For instance, the SICK data set contains examples like

T: Two young kids are eating corndogs.
H: Two young kids are fasting.
L: CONTRADICTION

Only with a definite article, *the two young kids*, T and H would be in contradiction. SNLI has plenty of similar examples:

> T: A couple walk hand in hand down the street.
> H: A couple is sitting on a bench.
> L: CONTRADICTION

There could be a situation where the T- H pairs are both true and there could be a picture showing the scene. What contradiction means in SICK and SNLI is that T and H are not captions for the same picture.

In a work on image descriptions Young et al. (2014: 68) develop a notion of visual denotations:

> "we propose to instantiate the abstract notions of possible worlds or situations with concrete sets of images. The interpretation function $[\![\cdot]\!]$ maps sentences to their **visual denotations** $[\![\mathbf{s}]\!]$ which is the set of images $\mathbf{i} \in U_s \subseteq U$ in a 'universe' of images U that **s** describes: $[\![\mathbf{s}]\!] = \{\ \mathbf{i} \in U \mid \mathbf{s}$ is a truthful description of $\mathbf{i}\}$"

In visual denotation semantics as presented in Young et al. (2014), a hypothesis H contradicts a premise T just in cases H is a caption for some image i that is not in the extension of the premise: $i \notin [\![\text{T}]\!]$.

The successor to SNLI is the MultiNLI (Williams et al., 2017) corpus presented at RepEval-2017 workshop. It consists of 433K examples. Unlike its immediate predecessors, MultiNLI is not based on image captions. Each seed sentence is picked from a broad range of written and spoken English. Each premise is derived from one of ten genre sections of the corpus. The hypotheses and the corresponding labels were generated by a crowd worker in response to a premise.

Because the premises in MultiNLI are not rephrased from image captions, the new data set is not skewed in the same bad way as SICK and SNLI are, where negative premises are exceedingly rare because of the way the premises we are derived. Image captions are typically about what is in the picture, not about things it doesn't show.

1.4 Our work on implicative constructions

The starting point for our research in deep learning technologies for textual inference was the long interest we have had in lexical semantics, in particular, *implicative constructions* (Karttunen, 1971; Karttunen and Peters, 1979; Karttunen, 2012, 2016). Implicative constructions include simple verbs that take infinitival complements such as *manage*, and *fail*, and a great number of verb+noun constructions such as *meet obligation* and *lack foresight*.

The characteristic feature of implicative constructions is that they yield an entailment about the the truth of the complement clause. *John managed to pass the exam* entails *John passed the exam*; *John did not manage to pass the exam* entails that he did not pass. Substituting *fail* for *manage* turns a positive entailment into a negative one, and vice versa.

Nairn et al. (2006) describes how the implicative inferences were computed in PARC's *Bridge* system in the early naughts. MacCartney (2009); MacCartney and Manning (2009) shows how the same inferences are obtained with MacCartney's *Natural Logic* rewrite rules.

Implicative constructions have not been systematically studied in the context of the studies discussed in the previous section. There are hardly any examples of them the RTE, SICK, and SNLI data sets. MultiNLI includes a scattering of examples with *manage* and *fail* but no examples of any

phrasal implicatives.[5] To address this shortcoming, we have assembled a large corpus and trained three simple models that learn implicatives.(Cases et al., 2017).

1.5 Goals for this workshop

In our previous work on implicatives we wanted to find out whether the vector representations of verb and noun meanings would enable our models to learn generalizations that are easy and intuitive for humans. For example, once you have learned that *Sally has met her obligation to file a tax return* entails that she has filed a tax return, you also know that if you replace *obligation* by a noun that has a similar meaning such as *duty*, *requirement* and *responsibility*, the new sentence has the same entailment as the original. The same goes for the verb *meet* in this construction. It could be replaced by *fulfill* and satisfy without affecting the entailment. Phrasal implicatives come in families. The experiments in Cases et al. (2017) showed that our model did poorly when it was tested completely novel constructions that had not been seen before in training or validation. But if some of the novel constructions were part of the validation set, the model showed, for example, that having been trained on examples with *miss chance*, *take chance* and *waste chance* it did perform well on analogous constructions with *opportunity*.

One problem related to the generalization capabilities is the relation between the implicative constructions themselves. Although the model has learned that *John solved the problem* is entailed by *John managed to solve the problem* and contradicted by *John failed to solve the problem* it is unable to classify the relation between *John managed to solve the problem* and *John failed to solve the problem* as a contradiction. As humans we know, without having been taught, that if A entails C and B entails ¬C, then A contradicts B, and vice versa.

We would like to construct models that not only learn simple one step inferences but also learn the basic properties of entailment and contradiction.

- Every statement entails itself.
- Entailment is transitive.
- Contradiction is symmetric.

We would like to hear your ideas about how to do that or be convinced that the problem is beyond what neural nets can do.

As the first step towards learning about transitivity, we have started a corpus of embedded implicatives. Implicatives constructions can be stacked. Here are some examples picked from the internet.

> Cindy had also managed`(+|-)` to forget`(-|+)`, to bring her makeup.
> I managed to remember to take pictures.
> I managed to fail to plug my phone in overnight.
> They failed to manage to grasp the reality of what was being said.
> It had failed to meet the duty to co-operate.
> The resident failed to have the wherewithal to respond to the unlawful detainer action.
> Freeland failed to obey the order to stop.
> I forgot to remember to forget her.

[5] MultiNLI contains 66 positive premises with the *manage to VP* construction but only 6 negative ones, probably too few to induce its inference pattern.

He also remembered to take time to celebrate.
Less than a fifth of the electorate bothered to turn out to vote.
No one bothered to take the trouble to ever tell you.
Millions of people dared to turn out to call for Mubarak to stand down.
A gull dared to have the gall to swoop down really low near a sailboat.

In the first example we have added the 'implicative signatures' of *manage*`(+|-)` and *forget*`(-|+)`, where `(+|-)` = positive entailment under positive polarity, negative entailment under negative polarity, `(-|+)` = negative entailment under positive polarity, positive entailment under negative polarity. Since the top-level sentence is positive, managed`(+|-)` preserves the positive polarity but the forget`(-|+)` clause flips the polarity from positive to negative for its complement. Thus the top level sentence entails that Cindy did not bring her makeup.

The RTE, SICK and SNLI contain no examples of phrasal implicatives. The MultiNLI corpus has a few examples of *take time* but probably too few to induce the correct `+|-` signature for this phrasal implicative because the examples are all positive. The model may learn that *the professor took the time to read my paper* entails *the professor read my paper* but, in the absence of any negative version of the premise, it will not learn that *the professor did not take the time to read my paper* entails that she did not do it.

There are hundreds of phrasal implicatives in English, most of them not well represented even in large data sets such as SNLI and MultiNLI. They are in the 'long tail' of the language that NLP systems eventually have to master to come close to the human performance. Our data set of implicative constructions is useful to that end even if it is 'synthetic,' that is, constructed by a program, by 'syntactic and lexical transformations with predictable effects' as was the case with SICK, instead of people as in SNLI and MultiNLI.

References

Marco Baroni and Roberto Zamparelli. 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*, pages 1183–1193. Association for Computational Linguistics, 2010. URL `http://aclanthology.coli.uni-saarland.de/pdf/D/D10/D10-1115.pdf`.

Daniel G. Bobrow, Bob Cheslow, Cleo Condoravdi, Lauri Karttunen, Tracy Holloway King, Rowan Nairn, Valeria de Paiva, Charlotte Price, and Annie Zaenen. PARC's bridge and question answering system. In Tracy Holloway King and Emily M. Bender, editors, *Proceedings of the GEAF07 Workshop*, pages 46–66, Stanford, CA, 2007. CSLI. URL `http://csli-publications.stanford.edu/GEAF/2007/geaf07-toc.html`.

Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 632–642. Association for Computational Linguistics, 2015.

Ignacio Cases, Lauri Karttunen, George Supaniratisai, and Arun Chaganty. An annotated corpurs of implicative constructions. Technical report, CSLI, 2017. Draft.

Qian Chen, Xiaodan Zhu, Zhenhua Ling, Si Wei, Hui Jiang, and Diana Inkpen. Enhanced LSTM for natural language inference. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2017.

Robin Cooper, Dick Crouch, Jan van Eijck, Chris Fox, Josef van Genabith, Jan Jaspars, Hans Kamp, David Milward, Manfred Pinkal, Massimo Poesio, and Steve Pulman. Using the framework. LRE 62-051 D-16. Technical report, University of Edinburgh, 1996. URL `http://www.cogsci.ed.ac.uk/~fracas/`.

Ido Dagan, Oren Glickman, and Bernardo Magnini. The PASCAL recognising textual entailment challenge. In *Proceedings of the First International Conference on Machine Learning Challenges: Evaluating Predictive Uncertainty Visual Object Classification, and Recognizing Textual Entailment*, MLCW'05, pages 177–190, Berlin, Heidelberg, 2006. Springer-Verlag. ISBN 3-540-33427-0, 978-3-540-33427-9. doi: 10.1007/11736790_9. URL `http://dx.doi.org/10.1007/11736790_9`.

Danilo Giampiccolo, Bernardo Magnini, Ido Dagan, and Bill Dolan. The third PASCAL recognizing textual entailment challenge, 2007. URL `http://aclanthology.coli.uni-saarland.de/pdf/W/W07/W07-1401.pdf`.

Edward Grefenstette and Mehrnoosh Sadrzadeh. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1394–1404. Association for Computational Linguistics, 2011. URL `http://aclanthology.coli.uni-saarland.de/pdf/D/D11/D11-1129.pdf`.

Andrew Hickl and Jeremy Bensley. A discourse commitment-based framework for recognizing textual entailment. In *Proceedings of the Workshop on Textual Entailment and Paraphrasing*, pages 171–176, 2007. URL `http://jmlr.org/proceedings/papers/v9/glorot10a/glorot10a.pdf`.

Lauri Karttunen. Implicative verbs. *Language*, 47(2):340–358, 1971. URL `\http://www.jstor.org/stable/i217134`.

Lauri Karttunen. Simple and phrasal implicatives. In **SEM 2012*, pages 124–131. Association for Computational Linguistics, Montréal, Canada, 7-8 June 2012. URL `http://www.aclweb.org/anthology/S12-1020`.

Lauri Karttunen. Presupposition: What went wrong? In Mary Moroney, Carol-Rose Little, Jacob Collard, and Dan Burgdorf, editors, *Semantics and Linguistic Theory (SALT) 26*, pages 705–731. Cornell University, Ithaca, NY, 2016. URL `http://web.stanford.edu/~laurik/publications/salt26.pdf`.

Lauri Karttunen and Stanley Peters. Conventional implicature. In Choon-Kyu Oh and David A. Dinneen, editors, *Syntax and Semantics, Volume 11: Presupposition*, pages 1–56. Academic Press, New York, 1979.

Bill MacCartney. *Natural language inference*. PhD thesis, Stanfor University, Stanford University, California, 2009.

Bill MacCartney and Christopher D. Manning. Natural logic for textual inference. In *Association for Computational Linguistics (ACL) Workshop on Textual Entailment and Paraphrasing*, 2007. URL `https://nlp.stanford.edu/pubs/natlog-wtep07.pdf`.

Bill MacCartney and Christopher D. Manning. An extended model of natural logic. In *The 8th International Conference on Computational Semantics (IWCS-8*, pages 140–156, Tilburg, Netherlands, 2009. University of Tilburg. URL `http://www.aclweb.org/anthology/W09-3700`.

Marco Marelli, Luisa Bentivogli, Marco Baroni, Raffaella Bernardi, Stefano Menini, and Roberto Zamparelli. 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, Dublin, Ireland, August 2014. Association for Computational Linguistics and Dublin City University. URL `http://www.aclweb.org/anthology/S14-2001`.

Jeff Mitchell and Mirella Lapata. Vector-based models of semantic composition. In *Proceedings of ACL-08: HLT*, pages 236–244. Association for Computational Linguistics, 2008. URL `http://aclanthology.coli.uni-saarland.de/pdf/P/P08/P08-1028.pdf`.

Rowan Nairn, Lauri Karttunen, and Cleo Condoravdi. Computing relative polarity for textual inference. In Johan Bos and Alexander Koller, editors, *Inference in Computational Semantics (ICoS-5)*, pages 67–76. University of Manchester, Manchester, UK, 2006. URL `http://www.aclweb.org/anthology/W06-39`.

Tim Rocktäschel, Edward Grefenstette, Karl Moritz Hermann, Tomás Kočiský, and Phil Blunsom. Reasoning about entailment with neural attention. *CoRR*, abs/1509.06664, 2015. URL `http://arxiv.org/abs/1509.06664`.

Adina Williams, Nikita Nangia, and Samuel R. Bowman. A broad-coverage challenge corpus for sentence understanding through inference. *CoRR*, abs/1704.05426, 2017. URL `http://arxiv.org/abs/1704.05426`.

Peter Young, Alice Lai, Micah Hodosh, and Julia Hockenmaier. From image descriptions to visual denotations. *Transactions of the Association for Computational Linguistics*, 2:67–78, 2014. URL `http://aclweb.org/anthology/Q/Q14/Q14-1006.pdf`.