A Distributional Theory of Content for NLP

Mark Steedman (*with* Mike Lewis)

1st June 2016

School of informatics

Outline

I: Paraphrase Cluster Semantics (Lewis and Steedman, 2013a)

II: Entailment-based Cluster Semantics (Lewis and Steedman, 2014b)

III: Multilingual Distributional Semantics (Lewis and Steedman, 2013b)

(?): Extension of the approach to Temporal Semantics

IV: Computational, Psychological, Linguistic, and Philosophical Conclusions.

The Problem of Content

- We have (somewhat) robust wide coverage parsers that work on the scale of Bn of words (e.g. Clark and Curran 2004; Lewis and Steedman 2014a). They can read the web (and build logical forms) thousands of times faster than we can ourselves.

- So why can't we ask them questions like "What are recordings by Miles Davis without Fender Rhodes piano", and get a helpful answer?

- The central problem of QA is that there are too many ways of asking and answering questions, and we have no idea of the semantics that relates them.

Too Many Ways of Answering The Question

- Your Question: *Did Google buy YouTube?*

- The Text:

  1. Google purchased YouTube.
  2. Google's purchase of YouTube
  3. Google acquired every company.
  4. YouTube may be sold to Google.
  5. Google will buy YouTube or Microsoft.
  6. Google didn't take over YouTube.

The Problem

- The hard problem in semantics is not the logical operators, but the content that they apply over.

- How do we define a theory of content that is robust in the sense of generalizing across linguistic form, and compositional in the sense of:
  - being compatible with logical operator semantics and
  - supporting commonsense inference?

Previous Work

- Many have tried and failed to build a form-independent semantics.

(1) Thomason, 1974: $\forall x[bug'x \Rightarrow \exists y[plants'(y) \wedge kill'y\ x]]$
McCawley, 1968: $[_S CAUSE\ BUGS[_S BECOME[_S NOT[_S ALIVE\ PLANTS]]]]$
Dowty, 1979: $[CAUSE[DO\ BUGS\ \varnothing][BECOME\neg[ALIVE\ PLANTS]]]$
Talmy, 2000: $Bugs\ ARE\text{-}the\text{-}AUTHOR''\text{-}OF[plants\ RESULT\text{-}TO\text{-}die]$
Van Valin, 2005: $[do'(bugs', \varnothing)]CAUSE[BECOME[dead'(plants)]]$
Goddard, 2010: *BUGS* do something to *PLANTS*; because of this, something happens to *PLANTS* at the same time; because of this, something happens to *PLANTS*s body; because of this, after this *PLANTS* are not living anymore.

Previous Work

- Cf. graphical representations of (Schank, 1972, Langacker, 2008, *passim*)

- Cf. WordNet (Fellbaum, 1998), VerbNet/PropBank (Hwang *et al.*, 2010), AMR (Banarescu *et al.*, 2012).

Hand-built semantic resources are inevitably incomplete.

- Why not let machine learning do the work instead?

- Treat the semantic primitives as hidden.

Two (Somewhat) New Approaches

- (Clustering by Collocation (Church and Hanks, 1989; Landauer and Dumais, 1997; Lin, 1998; Baroni and Zamparelli, 2010; Grefenstette and Sadrzadeh, 2011; Padó and Lapata, 2007; Mikolov *et al.*, 2013, *passim*).
  - Meanings are vectors
  - Composition is via Linear Algebraic Operations
  - Good for underspecification and disambiguation (Analogy tasks and Jeopardy questions).
- (Clustering by Denotation (Lin and Pantel, 2001; Hovy *et al.*, 2001; Lewis and Steedman, 2013a; Reddy *et al.*, 2014, *passim*).
  - Meanings are automatically extracted hidden relations.
  - Composition is via traditional Logical Operators
  - Good for inference of *entailment*.

I: Clustering by Paraphrase

- We seek to cluster expressions denoting the same relation. Instead of lexicons like the following:

(2) author $:= N/PP[of] : \lambda x \lambda y.author'xy$
write $:= (S\backslash NP)/NP : \lambda x \lambda y.write'xy$

- —we seek a lexicon capturing entailment via logical forms defined in terms of clusters of related meanings, like the following:

(3) author $:= N/PP_{of} : \lambda x_{book} \lambda y_{person}.relation37'xy$
write $:= (S\backslash NP)/NP : \lambda x_{book} \lambda y_{person}.relation37'xy$

- Such a "distributional" lexicon for content words works exactly like hand-built lexicons (1) with respect to the logical operator semantics of quantification and negation.

Method

- We obtain the clusters by parsing Gigaword text with the CCG-based C&C parser, augmented with the semantics from Steedman 2012, using a lexicon of the first type (2), to identify expressions relating Named Entities such as Google, YouTube, Scott, *Waverley*, etc.

- Nominal compounds for the same MUC named entity type are merged.

- Entities are soft-clustered into types according to a topic model based on LDA (Blei *et al.* (2003)) to induce type distributions for the named entities,

Method

Type	Most frequent members
1	suspect, assailant, fugitive, accomplice
2	author, singer, actress, actor, dad
5	city, area, country, region, town, capital
8	subsidiary, automaker, airline, Co., GM
10	musical, thriller, sequel, special

- The topic model types all words, not just named entity identifiers like GM.

Method

- These types are used to distinguish homonyms like the two versions of the *born in* relation relating PERSONS to DATES versus LOCATIONS
- Typed relations are hard-clustered based on Gigaword counts using a simple nonparametric algorithm Chinese Whispers (Biemann 2006; Fountain and Lapata 2011), which is highly scalable.
- Clustering is distributional, based on cosine similarities between tf-idf vectors of argument-pair counts for each predicate of a given type.
- We can then parse over full NPs in the target text using the clustered relations, as well as over the original named entities.

Example

- Obama was born in Hawaii.

$$(4)\quad \text{born} := (S\backslash NP)/PP[in] : \lambda x \lambda y. \left\{ \begin{array}{l} x = LOC \wedge y = PER \Rightarrow rel49 \\ x = DAT \wedge y = PER \Rightarrow rel53 \end{array} \right\} xy$$

$$\text{Obama} := \left\{ \begin{array}{l} PER = 0.9 \\ LOC = 0.1 \end{array} \right\}$$

$$\text{Hawaii} := \left\{ \begin{array}{l} LOC = 0.7 \\ DAT = 0.1 \end{array} \right\}$$

- The “Packed” Distributional Logical Form

$$(5)\quad S : \left\{ \begin{array}{l} rel49 = 0.63 \\ rel53 = 0.27 \end{array} \right\} hawaii' obama'$$

Evaluation: on Artificial Questions

- Our evaluation is based on Poon and Domingos (2009).
- We automatically construct a set of questions from answers found in dependency-parsed text, and then evaluate how many answers can be found in a different corpus.
- For example, from *Google bought YouTube*, we generate questions *What bought YouTube?* and *What did Google buy?*.
- We then attempt to answer the questions from a different text, same-genre corpus, using human judges to evaluate based on the sentence(s) found.
- Multiple answers count all answers.

Results: Question-Answer Test Set

- Examples:

Question	Answer	From Unseen Sentence:
What did Delta merge with?	Northwest	The 747 freighters came with Delta's acquisition of Northwest
What spoke with Hu Jintao?	Obama	Obama conveyed his respect for the Dalai Lama to China's president Hu Jintao during their first meeting
What arrived in Colorado?	Zazi	Zazi flew back to Colorado. . .
What ran for Congress?	Young	. . . Young was elected to Congress in 1972

- Full results in Lewis and Steedman (2013a)

Results: Question-Answer Test Suite

System	Answers	Correct
Relational LDA	7046	11.6%
REVERB	180	89.4%
CCG-Baseline	203	95.8%
CCG-WordNet	211	94.8%
CCG-Distributional@250	250	94.1%
CCG-Distributional@500	500	82.0%
- "Relational LDA" is Yao *et al.* 2011 trained on 35% of Gigaword.
- "REVERB is a sophisticated Open Information Extraction system (Fader *et al.*, 2011).

Fracas Test Suite

- Example:

Premises	Every European has the right to live in Europe Every European is a person Every person who has the right to live in Europe can travel freely within Europe
Hypothesis	Every European can travel freely within Europe
Solution:	Yes

- Further experiments including FRACAS in Lewis and Steedman 2013a.

II: Directional Entailments: The Hidden Language of Logical Form

- The above approach does not yet distinguish paraphrase from entailment.

- X_{person} *elected to* Y_{office} does entail X_{person} *ran for* Y_{office} but not *vice versa*.

- The paraphrase relation depends on more global properties of the named entity relation graph.

- Lewis (2015); Lewis and Steedman (2014b) apply the entailment graphs of Berant *et al.* (2012) to generate more articulated entailment structures.

Local Entailment Probabilities

- The typed named-entity technique is applied to (errorfully) estimate local probabilities of entailments using Weeds precision assymetric similarity (Weeds and Weir, 2003):

  a. $p(conquer\,x\,y \Rightarrow invade\,x\,y) = 0.9$

  b. $p(invade\,x\,y \Rightarrow attack\,x\,y) = 0.8$

  c. $p(conquer\,x\,y \Rightarrow attack\,x\,y) = 0.4$

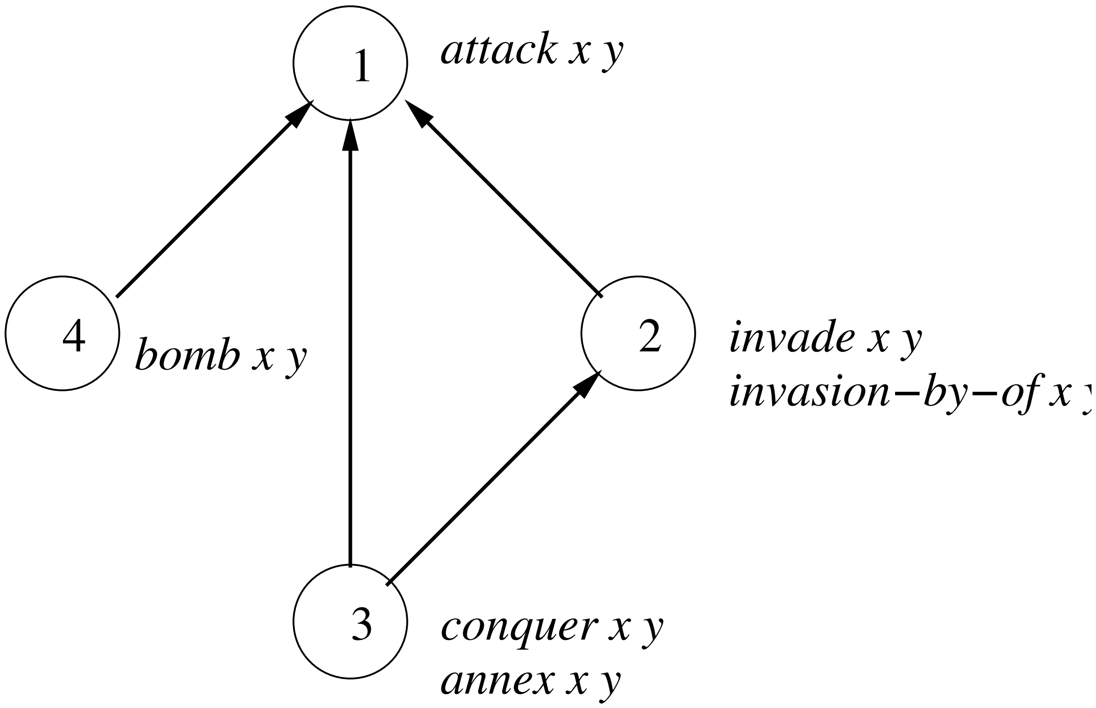
  d. $p(bomb\,x\,y \Rightarrow attack\,x\,y) = 0.7$

  (etc.)

Global Entailments

- The local entailment probabilities are used to construct an entailment graph using integer linear programming with a prior $p = 0.25$ with the global constraint that the graph must be closed under transitivity.
- Thus, (c) will be included despite low observed frequency, while other low frequency spurious local entailments will be excluded..
- Cliques within the entailment graphs are collapsed to a single paraphase cluster relation identifier, as in the previous approach.

Entailment graph



- A simple entailment graph for relations between countries.

Lexicon

- The lexicon obtained from the entailment graph

  attack $:= (S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1\, x\, y\, e$
  bomb $:= (S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1\, x\, y\, e \wedge rel_4\, x\, y\, e$
  invade $:= (S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1\, x\, y\, e \wedge rel_2\, x\, y\, e$
  conquer $:= (S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1\, x\, y\, e \wedge rel_2\, x\, y\, e \wedge rel_3\, x\, y\, e$
  annex $:= (S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1\, x\, y\, e \wedge rel_2\, x\, y\, e \wedge rel_3\, x\, y\, e$

- These logical forms support correct inference under negation, such as that *conquered* entails *attacked* and *didn't invade* entails *didn't conquer*
- To answer a question "Did X invade Y" we look for sentences which subsume the conjunctive logical form $rel_2 \wedge rel_1$, or satisfy its negation $\neg rel_2 \vee \neg rel_1$.
- Note that if we know that *invasion-of* is a paraphrase of $invade = rel_2$, we also know *invasion-of* entails $attack = rel_1$.

Experiment (Details—Skip)

- Train a local entailment classifier on a small entailment dataset of 5556 entailment problems based on pairs of Reverb extractions from Clueweb (Zeichner *et al.* 2012) parsed with C&C 50-best (10% of Zeichner is held out as a test set).
- For a Zeichner problem *Rome conquered Carthage* ⇒ *Rome invaded Carthage* we parse to make a training instance *conquer x y* ⇒ *invade x y*.
- We turn each training instance into a feature vector on which the classifier is trained as a function mapping vectors onto probabilities.
- The most important feature is argument pair distributional similarity in the larger Clueweb Reverb corpus. (E.g. distSim=0.3 for this positive instance. We use common noun heads as well as NEs.)
- The other features are derived by hand from the Zeichner training set e.g. morphological features, WordNet relations.

Experiment (Details—Skip)

- Next, for each ordered pairing of the n most common predicates in Clueweb we find their feature-vector representation, including typed NE distributional similarity in that corpus.
- We pass these to the Zeichner-supervised classifier, to obtain a probability that each represents an entailment.

Experiment (Details—Skip)

- Finally, we build an entailment graph on these most frequent relations in the (still too small) 15M dataset of Reverb propositions extracted from Clueweb, parsing with C&C.
- The graph includes the 100 most frequent relation expressions for 325 relation types such as PERSON+LOCATION
- Entity typing as Lewis and Steedman 2013a (25 topics).
- A generalization of Berant *et al.* 2012 using the Zeichner set as well as Wordnet relations etc. for the local classifier.
- Evaluate over held out Zeichner entailment data as test set by parsing the sentences into packed logical forms including negation and quantifiers (Steedman, 2012), using the Prover9 theorem prover.

Experiment: Evaluation

- Testset is held-out data from the Zeichner *et al.* (2012) entailment set.
- Baselines are Majority Class (don't know) and Berant *et al.* 2011 Non Compositional direct entailment between reverb patterns.
- We also compare with Additive and Multiplicative Vector-based distributional semantics (SCS) using a logistic regression classifier.
- The Zeichner entailments, unlike RTE, rely predominantly on lexical entailment.

- This dataset does not otherwise play to the syntactic and logical strengths of CCG, and includes many non-compositional idioms (eg light verb construction) quite favorable to e.g. vector composition.
- Zeichner has No negation. No quantifiers. :@(

Examples from Zeichner *et al.*, 2012

Premise	Hypothesis	Answer
Obama want to boost the defense budget	Obama increase the defense budget	False
The thieves make off with TVs	The thieves manage to steal TVs	True
My son be terrified of him	My son have a fear of him	True

Results

System	**Accuracy (all)**	**AUC (all)**
Majority Class	56.8%	0.46
Non Compositional	57.4%	0.48
CCG Baseline	57.8%	0.46
Lewis and Steedman (2013a)	58.0%	0.50
VectorMultiplicative	61.3%	0.51
VectorAdditive	63.5%	0.57
CCG Entailment Graphs	64.9%	0.61
CCG Entailment Graphs+ Implicative Verb Lexicon	**66.0%**	**0.62**

- Last line shows the effect of adding 50 hand-coded frequent implicative verbs where *managing to win* entails *winning*, while *failing to win* entails *not winning* (Bos, 2013).

- AUC is area under Precision-Recall curve, computed with a trapezoid approximation, as a measure of reliability of confidence estimates.

III: Clustering Cross-linguistically

- We use cross-linguistic paraphrase clusters to re-rank Moses n-best lists to promote translations that preserve the cluster-based meaning representation from source to target.

- This requires a reasonably accurate parser for the source and target languages—not necessarily CCG based. . .

- . . . although CCG helps—see Boonkwan and Steedman (2011); Boonkwan (2013) and Ambati *et al.* (2013, 2014) on parsing under-resourced languages.

Experiment: Reranking Moses Translations

- For a source (French) sentence that can be dependency-parsed to deliver a cluster-semantic logical form:
- We Moses-translate (to English) taking the 50-best list and parsing (with C&C) to produce cluster-semantic logical forms.
- If the logical form of the top ranked translation is the same as that of the source sentence, we discard this trial as uninformative.
- If the logical form of the top ranked translation is different from the source, we choose whatever translation from the remainder of the n-best list has the logical form that most closely resembles the source cluster semantics.
- Fluent bilingual human annotators are then asked to choose between the one-best Moses translation and the cluster-based alternative.

Example

Source:	Le Princess Elizabeth arrive à Dunkerque le 3 août 1999
SMT 1-best:	The Princess Elizabeth is to manage to Dunkirk on 3 August 1999.
Reranked 1-best:	The Princess Elizabeth arrives at Dunkirk on 3 August 1999.

Reranking Moses

-

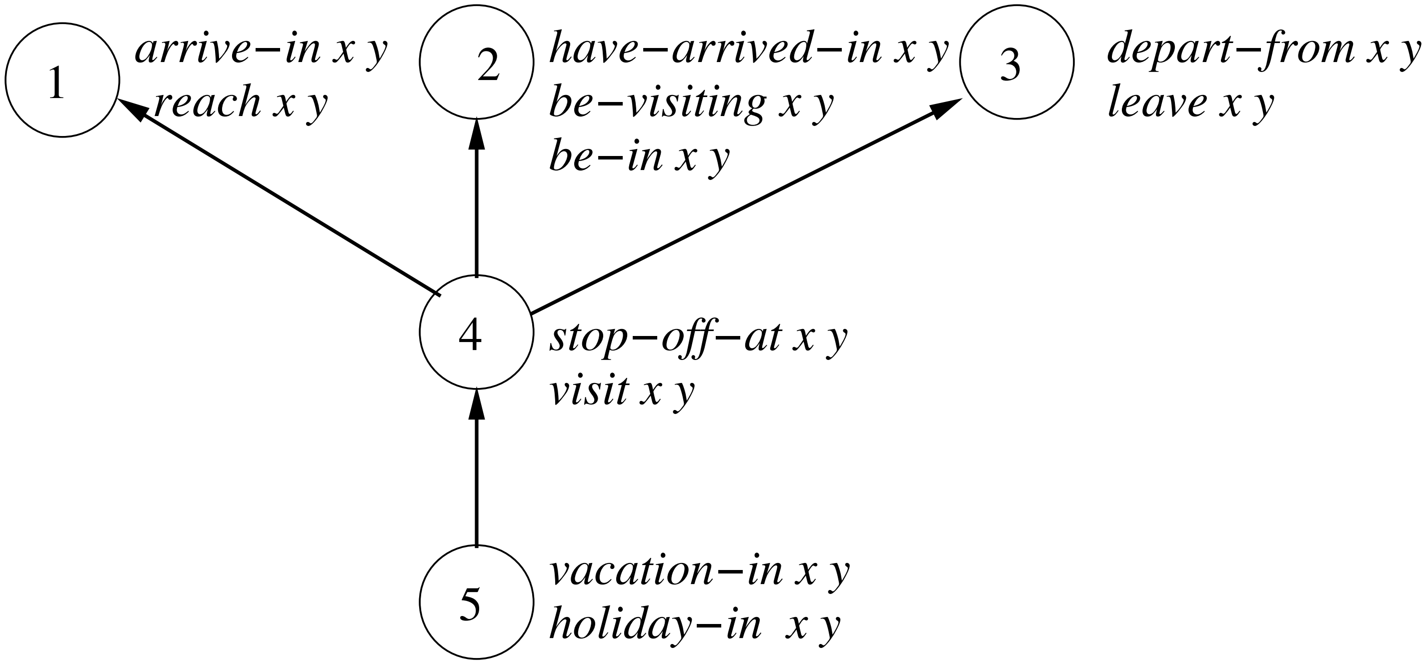
	Percentage of Translations preferred
1-best Moses	5%
Reranked best	39%
No preference	56%

- Many cases of "no preference" were where Moses and the prefered translation were similar strings differing only in attachment decisons invisible to the human judges.
- No parallel text was harmed/used in these experiments.
- This is good, because SMT has already used up all of the available parallel text (Och, 2007)!
- Full results in Lewis and Steedman (2013b).

IV: What Relations Can We Learn This Way?

- The most urgent extension needed is to one place relations, many of which are nominal. This would amount to automatically building or extending WordNet using the present technique.
- The strong effect of our hand-coded implicative verbs like “X managed to Y” as entailing “X Yed” suggests that it would be possible to learn entailment graphs over them and their paraphrases in the same way as main verb relations.
- The same observation applies to light verb constructions, like “Take a trip”.
- Presuppositions which are entailed both by Factive verbs like “know” and their negation, are treated non-conjunctively, as arising from factive definite reference.

Generalizing Entailment to Temporal Semantics



- A simple entailment graph for relations over events does not capture relations of causation and temporal sequence entailment.

Temporal Semantics

- As in the case of the semantics of content words like nouns and verbs, the semantics of tense, aspect, modality, evidentiality, and intensionality has always seemed to bog down in conflicting and overlapping ontology, and ill-defined or world-knowledge-entangled notions like "inertia worlds", "relevance", "extended now", "perfect time span", "consequent state", "preparatory activity", and the like.
  - #Einstein has visited New York (*vs.* Einstein visited New York).
  - #I have forgotten your name but I have remembered it again (*vs.* I forgot your name but I remembered it again).
- Such relations seem like A Suitable Case for Treatment as hidden relations, letting machine learning find out what the consequent states of people *visiting* places, *forgetting* and *remembering* things, etc. usually are.

Temporal Semantics from Timestamped Data

- Pilot experiments have begun with timestamped news under a Google Faculty Award, using the University of Washington NewsSpike corpus of 0.5M newswire articles (Zhang and Weld, 2013).

{"arg1":"OBAMA","arg2":"MINNEAPOLIS","sentences":
[{"relationphrase":"be in","tokens":
["Obama","is","in","Minneapolis","to","push","for","tougher","gun","laws","and","highlight","some","of","the","things",
"a1":[0,1],"a2":[3,4],"v":[1,3],"fromArticleId":371037},
{"relationphrase":"head to","tokens":
["Obama","heads","to","Minneapolis","to","sell","gun","plan","."],
"a1":[0,1],"a2":[3,4],"v":[1,3],"fromArticleId":369952},
{"relationphrase":"be visit","tokens":
["Monday",",","Obama","is","visiting","Minneapolis","to","discuss","his","plan","to","battle","gun","violence","."],
"a1":[2,3],"a2":[5,6],"v":[3,5],"fromArticleId":433846}], ... }
{"arg1":"DAVID BECKHAM","arg2":"PARIS","sentences":
[{"relationphrase":"have arrive
in","tokens":["David","Beckham","has","arrived","in","Paris","to","complete","a","dramatic","deadline","day","move","to
"a1":[0,2],"a2":[5,6],"v":[2,5],"fromArticleId":456691},
{"relationphrase":"go
to","tokens":["David","Beckham","Goes","to","Paris",",","Kate","Middleton","Shops","Incognito",",","and","Dolce","\u002
"a1":[0,2],"a2":[4,5],"v":[2,4],"fromArticleId":452413}], ... }

Timestamped Data

- In such data, we find that statements that so-and-so *is visiting*, *is in* and the perfect *has arrived in* such and such a place, occur in stories with the same datestamp, whereas *is arriving*, *is on her way to*, occur in preceding stories, while *has left, is on her way back from, returned*, etc. occur in later ones.
- This information provides a basis for inference that *visiting* entails *being in*, that the latter is the consequent state of *arriving*, and that *arrival* and *departure* coincide with the beginning and end of the progressive state of *visiting*.

Reconnecting with Logical Operator Semantics

- Some handbuilt lexical entries for auxiliary verbs (closed-class words):

  has $:= (S\backslash NP)/(S_{en}\backslash NP) : \lambda p_E \lambda y.\mathit{consequent\text{-}state}(p_E\, y)\,\mathbf{R} \wedge \mathbf{R} = \mathbf{NOW}$

  will $:= (S\backslash NP)/(S_b\backslash NP) : \lambda p_E \lambda y.\mathit{priors} \Rightarrow \mathit{imminent\text{-}state}(p_E\, y)\,\mathbf{R})$
  $\wedge\, \mathbf{R} = \mathbf{NOW}$

  is $:= (S\backslash NP)/(S_{ing}\backslash NP) : \lambda p_E \lambda y.\mathit{progressive\text{-}state}(p_E\, y)\,\mathbf{R} \wedge \mathbf{R} = \mathbf{NOW}$

- Cf. Steedman, 1977; Webber, 1978; Steedman, 1982; Moens and Steedman, 1988; White, 1994; Steedman, 1997.

Reconnecting with Logical Operator Semantics

- Some potentially learnable lexical entries for implicative verbs:

$$\text{tried} := (S\backslash NP)/(S_{to}\backslash NP) : \lambda p_E \lambda y.rel_{try}\, p_E\, y\, \mathbf{R} \wedge rel_{want}\, p_E\, y\, \mathbf{R}$$
$$\wedge\, preparatory\text{-}activity\,(p_E\, y)\, y\, \mathbf{R} \wedge \mathbf{R} < \mathbf{NOW}$$

$$\text{failed} := (S\backslash NP)/(S_{to}\backslash NP) : \lambda p_E \lambda y.rel_{try}\, p_E\, y\, \mathbf{R} \wedge rel_{want}\, p_E\, y\, \mathbf{R}$$
$$\wedge\, preparatory\text{-}activity\,(p_E\, y)\, y\, \mathbf{R} \wedge \neg p_E\, y\, \mathbf{R} \wedge \mathbf{R} < \mathbf{NOW}$$

$$\text{managed} := (S\backslash NP)/(S_{to}\backslash NP) : \lambda p_E \lambda y.rel_{try}\, p_E\, y\, \mathbf{R} \wedge rel_{want}\, p_E\, y\, \mathbf{R}$$
$$\wedge\, preparatory\text{-}activity\,(p_E\, y)\, y\, \mathbf{R} \wedge p_E\, y\, \mathbf{R} \wedge \mathbf{R} < \mathbf{NOW}$$

Conclusions I: For Computational Linguistics

- Learning over denotations of relations over typed named entities allows us to construct logical forms for content words as distributions over typed conjunctions of entailments over paraphrase clusters.
- These conjunctive terms in this logical language are very close to the language-specific grammar, and support fast inference of common-sense entailment.
- Under more traditional semantic theories employing eliminative definitions these entailments would have been thought of as belonging to the domain of inference rather than semantics, either as meaning postulates relating logical forms or as "encyclopædic" general knowledge.
- This meaning representation is compatible with a traditional logical operator semantics.

Conclusions II: For Philosophy of Language

- Carnap (1952) introduced meaning postulates in support of Inductive Logic, including a model of Probability, basically to keep the model small and consistent.
- Like Katz and Fodor (1963); Katz and Postal (1964); Katz (1971), we are in effect packing meaning postulates into the lexicon.
- This suggests that our semantic representation expresses an a pragmatic empiricist view of analytic meaning of the kind advocated by Quine (1951).
- It can also be viewed as a grammar-based statistical model of "meaning as use" (Wittgenstein, 1953).

Conclusions III: For Psychology

- Do children acquire the meaning of words like "annex" and "conquer" by building entailment graphs?
- I suggest they do, and that this is the mechanism for what Gleitman (1990) called syntactic bootstrapping of the lexicon—that is:
  - Once children have acquired core competence (by semantic bootstrapping of the kind modeled computationally by Kwiatkowski *et al.* 2012 and Abend *et al.*, 2016), they can detect that "annex" is a transitive verb meaning some kind of attack without knowing exactly what it means.
  - They can then acquire the full meaning by piecemeal observation of its entailments and paraphrases in use.
  - This is a major mechanism of cultural inheritance of concepts that would otherwise in many cases take more than an individual lifetime to develop.

Conclusions IV: For Cognitive Science

- These terms compile into a (still) language-specific Language of Thought (Fodor 1975, *passim*), which is roughly what adult speakers do their thinking in.
- To the extent that the cliques or clusters in the graph are constructed from multilingual text, this meaning representation will approximate the hidden language-independent "private" Language of Mind which the child language learner accesses.
- However, very few terms in any adult logical form correspond directly to the hidden primitives of that Language of Mind. (*red* and maybe *attack* might be exceptions.)
- Even those terms that are cognitively primitive (such as color terms) will not be unambiguously lexicalized in all languages.

Conclusions V: For Artificial Intelligence

- Some conceptual primitives, such as that things can only be in one place at a time, probably predate human cognition, and are unlikely to be discoverable at all by machine reading of the kind advocated here.
- These properties are hard-wired into our minds by 600M years of vertebrate evolution.
- These are exactly the properties that Artificial Intelligence planning builds in to the representation via the "Closed World Assumption" and the STRIPS dynamic logic of change.
- Computational Lingustics should learn from AI in defining a Linear Dynamic Logic for distributional clustered entailment semantics.

References

Abend, Omri, Kwiatkowski, Tom, Smith, Nathaniel, Goldwater, Sharon, and Steedman, Mark, 2016. "Bootstrapping Language Acquisition." *submitted* :1–40.

Ambati, Bharat Ram, Deoskar, Tejaswini, and Steedman, Mark, 2013. "Using CCG Categories to Improve Hindi Dependency Parsing." In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Sofia, 604–609.

Ambati, Bharat Ram, Deoskar, Tejaswini, and Steedman, Mark, 2014. "Improving Dependency Parsers using Combinatory Categorial Grammar." In *Proceedings* *of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*. Gothenburg, 159–163.

Banarescu, Laura, Bonial, Claire, Cai, Shu, Georgescu, Madalina, Griffitt, Kira, Hermjakob, Ulf, Knight, Kevin, Koehn, Philipp, Palmer, Martha, and Schneider, Nathan, 2012. "Abstract Meaning Representation (AMR) 1.0 Specification."

Baroni, Marco and Zamparelli, Roberto, 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*. Cambridge, MA: ACL, 1183–1193.

Berant, Jonathan, Dagan, Ido, Adler, Meni, and Goldberger, Jacob, 2012. "Efficient Tree-Based Approximation for Entailment Graph Learning." In *Proceedings of the 50th Annual Meeting of the Association for Computational*

Linguistics: Long Papers-Volume 1. Association for Computational Linguistics, 117–125.

Berant, Jonathan, Goldberger, Jacob, and Dagan, Ido, 2011. “Global Learning of Typed Entailment Rules.” In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. Portland, OR, 610–619.

Biemann, Chris, 2006. “Chinese Whispers: An Efficient Graph Clustering Algorithm and its Application to Natural Language Processing Problems.” In *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*. ACL, 73–80.

Blei, David, Ng, Andrew, and Jordan, Michael, 2003. “Latent Dirichlet Allocation.” *Journal of Machine Learning* 3:993–1022.

Boonkwan, Prachya, 2013. *Scalable Semi-Supervised Grammar Induction* *using Cross-Linguistically Parameterized Syntactic Prototypes*. Ph.D. thesis, Edinburgh.

Boonkwan, Prachya and Steedman, Mark, 2011. "Grammar Induction from Text Using Small Syntactic Prototypes." In *Proceedings of the International Joint Conference on Natural Language Processing*. Bangkok: ACL, 438–446.

Bos, Johan, 2013. "Is There a Place for Logic in Recognizing Textual Entailment?" *Linguistic Issues in Language Technology* 9.

Carnap, Rudolf, 1952. "Meaning Postulates." *Philosophical Studies* 3:65–73. reprinted as Carnap, 1956:222-229.

Carnap, Rudolf (ed.), 1956. *Meaning and Necessity*. Chicago: University of Chicago Press, second edition.

Church, Kenneth and Hanks, Patrick, 1989. "Word Association Norms, Mutual Information, and Lexicography." In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*. ACL, 76–83.

Clark, Stephen and Curran, James R., 2004. "Parsing the WSJ using CCG and Log-Linear Models." In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*. Barcelona: ACL, 104–111.

Dowty, David, 1979. *Word Meaning in Montague Grammar*. Dordrecht: Reidel, first edition.

Fader, Anthony, Soderland, Stephen, and Etzioni, Oren, 2011. "Identifying Relations for Open Information Extraction." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL, 1535–545.

Fellbaum, Christiane (ed.), 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Fodor, Jerry, 1975. *The Language of Thought*. Cambridge, MA: Harvard.

Fountain, Trevor and Lapata, Mirella, 2011. "Incremental Models of Natural Language Category Acquisition." In *Proceedings of the 32st Annual Conference of the Cognitive Science Society*. 255–260.

Gleitman, Lila, 1990. "The Structural Sources of Verb Meanings." *Language Acquisition* 1:1–55.

Goddard, Cliff, 2010. "The Natural Semantic Metalanguage Approach." In Bernd Heine and Heiko Narro (eds.), *The Oxford Handbook of Linguistic Analysis*, Oxford: Oxford University Press. 459–484.

Grefenstette, Edward and Sadrzadeh, Mehrnoosh, 2011. “Experimental Support for a Categorical Compositional Distributional Model of Meaning.” In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Edinburgh: ACL, 1394–1404.

Hovy, Eduard, Gerber, Laurie, Hermjakob, Ulf, Junk, Michael, and Lin, Chin-Yew, 2001. “Question Answering in Webclopedia.” In *Proceedings of the Ninth Text Retrieval Conference (TREC-9)*. Washington, DC: NIST, 655–664.

Hwang, Jena D, Bhatia, Archna, Bonial, Clare, Mansouri, Aous, Vaidya, Ashwini, Xue, Nianwen, and Palmer, Martha, 2010. “Propbank annotation of multilingual light verb constructions.” In *Proceedings of the Fourth Linguistic Annotation Workshop*. Association for Computational Linguistics, 82–90.

Katz, Jerrold and Fodor, Jerry, 1963. “The Structure of a Semantic Theory.” *Language* :170–210.

Katz, Jerrold and Postal, Paul, 1964. *An Integrated Theory of Linguistic Descriptions*. Cambridge, MA: MIT Press.

Katz, Jerrold J, 1971. "Generative Semantics is Interpretive Semantics." *Linguistic inquiry* 2(3):313–331.

Kwiatkowski, Tom, Goldwater, Sharon, Zettlemoyer, Luke, and Steedman, Mark, 2012. "A Probabilistic Model of Syntactic and Semantic Acquisition from Child-Directed Utterances and their Meanings." In *Proceedings of the 13th Conference of the European Chapter of the ACL (EACL 2012)*. Avignon: ACL, 234–244.

Landauer, Thomas and Dumais, S., 1997. "A Solution to Plato's Problem: The Latent Semantic Analysis of the Acquisition, Induction and Representation of Knowledge." *Psychological Review* 104:211–240.

Langacker, Ronald, 2008. *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

Lewis, Mike, 2015. *Combined Distributional and Logical Semantics*. Ph.D. thesis, University of Edinburgh.

Lewis, Mike and Steedman, Mark, 2013a. "Combined Distributional and Logical Semantics." *Transactions of the Association for Computational Linguistics* 1:179–192.

Lewis, Mike and Steedman, Mark, 2013b. "Unsupervised Induction of Cross-Lingual Semantic Relations." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL, 681–692.

Lewis, Mike and Steedman, Mark, 2014a. "A^* CCG Parsing with a Supertagfactored Model." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Doha, Qatar: ACL, 990–1000.

Lewis, Mike and Steedman, Mark, 2014b. "Combining Formal and Distributional Models of Temporal and Intensional Semantics." In *Proceedings of the ACL Workshop on Semantic Parsing*. Baltimore, MD: ACL, 28–32. Google Exceptional Submission Award.

Lin, Dekang, 1998. "Automatic Retrieval and Clustering of Similar Words." In *Proceedings of the International Conference on Computational Linguistics/Annual Meeting of the Association for Computational Linguistics*. Montreal, 768–774.

Lin, Dekang and Pantel, Patrick, 2001. "DIRT—Discovery of Inference Rules from Text." In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data-Mining (KDD-01)*. San Francisco, 323 – 328.

McCawley, James, 1968. "Lexical Insertion in a Transformational Grammar without Deep Structure." In *Papers from the 4th Regional Meeting of the Chicago Linguistics Society*. CLS, 71–80.

Mikolov, Tomas, Sutskever, Ilya, Chen, Kai, Corrado, Greg, and Dean, Jeff, 2013. "Distributed Representations of Words and Phrases and their Compositionality." In *Advances in Neural Information Processing Systems*. 3111–3119.

Moens, Marc and Steedman, Mark, 1988. "Temporal Ontology and Temporal Reference." *Computational Linguistics* 14:15–28. reprinted in Inderjeet Mani, James Pustejovsky, and Robert Gaizauskas (eds.) *The Language of Time: A Reader*. Oxford University Press, 93-114.

Och, Franz, 2007. "Keynote." In *Workshop on Statistical Machine Translation*. ACL.

Padó, Sebastian and Lapata, Mirella, 2007. "Dependency-Based Construction of Semantic Space Models." *Computational Linguistics* 33:161–199.

Quine, Willard Van Orman, 1951. "Two dogmas of empiricism." *The Philosophical Review* :20–43reprinted in Quine (1953).

Quine, Willard Van Orman, 1953. *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

Reddy, Siva, Lapata, Mirella, and Steedman, Mark, 2014. "Large-scale Semantic Parsing without Question-Answer Pairs." *Transactions of the Association for Computational Linguistics* 2:377–392.

Schank, Roger, 1972. "Conceptual Dependency: A Theory of Natural Language Understanding." *Cognitive Psychology* 3:552–631.

Steedman, Mark, 1977. "Verbs, Time and Modality." *Cognitive Science* 1:216–234.

Steedman, Mark, 1982. "Reference to Past Time." In Robert Jarvella and Wolfgang Klein (eds.), *Speech, Place, and Action*, New York: Wiley. 125–157.

Steedman, Mark, 1997. "Temporality." In Johan van Benthem and Alice ter Meulen (eds.), *Handbook of Logic and Language*, Amsterdam: North Holland/Elsevier. 895–938.

Steedman, Mark, 2000. *The Syntactic Process*. Cambridge, MA: MIT Press.

Steedman, Mark, 2012. *Taking Scope: The Natural Semantics of Quantifiers*. Cambridge, MA: MIT Press.

Talmy, Leonard, 2000. *Towards a Cognitive Semantics*, volume 1 and 2. Cambridge, MA: MIT Press.

Thomason, Richmond (ed.), 1974. *Formal Philosophy: Papers of Richard Montague*. New Haven, CT: Yale University Press.

Van Valin, Robert (ed.), 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.

Webber, Bonnie, 1978. "On Deriving Aspectual Sense: a Reply to Steedman." *Cognitive Science* 2:385–390.

Weeds, Julie and Weir, David, 2003. "A General Framework for Distributional Similarity." In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*. ACL, 81–88.

White, Michael, 1994. *A Computational Approach to Aspectual Composition*. Ph.D. thesis, University of Pennsylvania.

Wittgenstein, Ludwig, 1953. *Philosophical Investigations*. London: Basil Blackwell.

Yao, Limin, Haghighi, Aria, Riedel, Sebastian, and McCallum, Andrew, 2011. "Structured Relation Discovery using Generative Models." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL, Edinburgh, 1456–1466.

Zeichner, Naomi, Berant, Jonathan, and Dagan, Ido, 2012. "Crowdsourcing Inference-Rule Evaluation." In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*. ACL, 156–160.

Zhang, Congle and Weld, Daniel, 2013. "Harvesting Parallel News Streams to Generate Paraphrases of Event Relations." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Seattle: ACL, 1776–1786.