A Corpus and Model Integrating Multiword Expressions and Supersenses

Nathan Schneider
Noah Smith

NAACL-HLT • June 3, 2015, Denver
Given a sentence
find & categorize
minimal units of meaning
cheaply, with broad coverage
Noam Chomsky refused to give in to the vicious daddy longlegs.
Noam Chomsky refused to give in to the vicious daddy longlegs.
Noam Chomsky refused to give in to the vicious daddy longlegs.
Noam Chomsky refused to give in to the vicious daddy longlegs.
Lexical segmentation

Noam_Chomsky refused to give_in_to the vicious daddy_longlegs.
give_in_to
daddy_longlegs
Noam_Chomsky
Noam Chomsky refused to give in to the vicious daddy longlegs.
Outline

- Background
  - multiword expressions
  - supersenses
- Dataset
- Joint model
- Results
Definition
(Baldwin & Kim, 2010; Schneider et al., LREC 2014)

- Multiword expression (MWE): 2 or more orthographic words/lexemes that function together as an **idiomatic whole**

- **idiomatic** = not fully predictable in form, function, and/or frequency
  - unusual morphosyntax: *Me/*Him neither; **by and large**; plural of **daddy longlegs**?
  - non- or semi-compositional: **ice cream**, **daddy longlegs**, **pay attention**
  - statistically collocated: $p(\text{highly unlikely}) > p(\text{strongly unlikely})$
Definition
(Baldwin & Kim, 2010; Schneider et al., LREC 2014)

- **Multiword expression** (MWE): 2 or more orthographic words/lexemes that function together as an **idiomatic whole**

- *idiomatic* = not fully predictable in form, function, and/or frequency

  - unusual morphosyntax: Me/*Him neither; by and large, plural of daddy longlegs?
  - non- or semi-compositional: ice cream, daddy longlegs, pay attention
  - statistically collocated: \( p(\text{highly unlikely}) > p(\text{strongly unlikely}) \)

SPECIALLY LEARNED
Noam Chomsky

daddy longlegs, hot dog

dry out the clothes

depend on, come across

put up with, give in (to)

under the weather

cut and dry

in spite of

pick up where they left off

easy as pie

You’re welcome.

To each his own.

The structure of this paper is as follows.
The CMWE Corpus

(Schneider et al., LREC 2014)

- The entire **REVIEWS** subsection of the English Web Treebank (Bies et al. 2012), comprehensively annotated for MWEs
  - 723 reviews
  - 3,800 sentences
  - 55,000 words
  - found 3,500 MWE instances
  - 57% of all sentences (72% >10 words) contain an MWE
They gave me the run around and missing paperwork only to call back to tell me someone else wanted her and I would need to come in and put down a deposit.
CMWE Example
(Schneider et al., LREC 2014)

They *gave* me *the_run_around* and missing paperwork only to *call_back* to tell me someone else wanted her and I would need to *come_in* and *put_down* a deposit.

Simplified a bit for presentational purposes (we also made a strong/weak distinction)
I'm in the green room getting ready for my panel #textworld

V:STATIVE
N:LOCATION
V:COGNITION
N:GROUP

supersenses
Supersenses

- Semantic classes originally defined by WordNet
- Can be inferred from WordNet annotations in SemCor (Miller et al. 1993)
- ...or annotated directly (Schneider et al. 2012: Arabic Wikipedia; this work)
  - also Johannsen et al. 2014: English Twitter
- Automatic tagging (Ciaramita & Altun 2006; Paaß & Reichartz 2009; Schneider et al. 2013; Johannsen et al. 2014)
Outline

✓ Background
  ‣ multiword expressions
  ‣ supersenses

• Dataset

• Joint model

• Results
STREUSLE Corpus

Supersense Tagged Repository of English with a Unified Semantics for Lexical Expressions
STREUSLE Corpus

- Annotated with
  - comprehensive MWEs
  - noun+verb supersenses
I googled restaurants in the area and Fuji_Sushi came_up and reviews were great so I made a carry_out order
I googled restaurants in the area and Fuji Sushi came up and reviews were great so I made a carry out order.
I googled restaurants in the area and Fuji_Sushi came up and reviews were great so I made a carry_out order
I googled restaurants in the area and Fuji_Sushi came up and reviews were great so I made a carry_out order
I googled restaurants in the area and Fuji_Sushi came_up and reviews were great so I made_ a carry_out_order
I googled restaurants in the area and Fuji_Sushi came_up and reviews were great so I made_ a carry_out _order
STREUSLE Annotation

- **Starting point:** CMWE corpus
- **2 main phases:**
  - noun supersenses
  - verb supersenses
- Some sentences were reserved for combined **noun+verb** annotation
STREUSLE Annotation

- Preexisting conventions for **noun** supersenses that were applied to Arabic Wikipedia (Schneider et al., 2012)

- **This work**: New conventions for **verb** supersenses
STREUSLE Annotation: Verbs

cognition (thinking, judging, analyzing, doubting)
  decide, think, rate (assign rating), respect = have respect for, memorize, learn, see = understand
  contrast with perception, communication

communication (verbal/linguistic or nonverbal gesturing: telling, asking, ordering)
  speak, talk, write = communicate by writing, announce, type (on a keyboard), cry out, describe, argue, contest, petition, stammer, beg, mandate, veto, libel, preach, teach (education), fax, moo (animal noise)
  • WN lists music production (a person singing/playing an instrument) as creation
  • noises from inanimate objects (‘creak’, etc.) are perception
  • contrast with perception, cognition

competition (fighting, athletic activities)
  compete, fight (with someone), play (sports), referee, duel [supersedes social?; superseded by communication for rhetorical senses of ‘attack’, ‘contend’, etc.; superseded by contact for moments of physical contact: ‘wrestle’, ‘box’, ‘punch’, ‘beat up’]
STREUSLE Annotation: Verbs

Precedence relations

- \{ perception, consumption \} > body > change
- motion > social > change
- emotion > change
- motion > \{ body, possession \} (e.g., stand_up, bring)
- contact > \{ stative, motion \}
- \{ contact, communication \} > competition > social
- emotion > cognition
STREUSLE IAA

• We estimated inter-annotator $F_1$ of supersense labels at the end of each phase of annotation.
  ‣ Nouns-only phase: 76%
  ‣ Verbs-only phase: 93%
  ‣ Combined phase: 88%
Outline

✓ Background
  ▸ multiword expressions
  ▸ supersenses
✓ Dataset
  ● Joint model
  ● Results
Outline

✓ Background
  ‣ multiword expressions
  ‣ supersenses

✓ Dataset
  ● Joint model

● Results
Gappy sequence tagging
(Schneider et al., TACL 2014)

• *Contiguous* MWE identification resembles chunking, so we can use the familiar BIO scheme (Ramshaw & Marcus 1995):

```
O O B I O
a routine oil_change .
```

• 3 new tags for *gaps*:

```
0 0 0 B o b i i i I
My wife had taken_ her '07_Ford_Fusion _in
```

- Assumption: no more than 1 level of nesting

• **Evaluation**: MWE precision/recall
  - Link-based: partial credit for partial overlap
Gappy sequence tagging
(Schneider et al., TACL 2014)

- Standard supervised learning with the enriched tagging scheme
- **Structured perceptron** (Collins 2002)
  - Discriminative
  - 1st-order Markov assumption
  - Averaging
  - Fast to train
Gappy sequence tagging
(Schneider et al., TACL 2014)

- **Basic features**
  adapted from Constant et al. (2012):
  - **word:** current & context, unigrams & bigrams
  - **POS:** current & context, unigrams & bigrams
  - capitalization; word shape
  - prefixes, suffixes up to 4 characters
  - has digit; non-alphanumerical characters
  - lemma + context lemma if one is a V and the other is ∈ {N, V, Adj., Adv., Prep., Part.}

- **Lexicon features:** WordNet & other lexicons
Joint Tag Encoding

- Augment the MWE tags with supersense labels

<table>
<thead>
<tr>
<th></th>
<th>MWE only</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>My</strong></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>wife</strong></td>
<td>0</td>
<td>0-PERSON</td>
</tr>
<tr>
<td><strong>had</strong></td>
<td>0</td>
<td>0-`a</td>
</tr>
<tr>
<td><strong>taken</strong></td>
<td>B</td>
<td>B-motion</td>
</tr>
<tr>
<td><strong>her</strong></td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td><strong>'07</strong></td>
<td>b</td>
<td>b-ARTIFACT</td>
</tr>
<tr>
<td><strong>Ford</strong></td>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td><strong>Fusion</strong></td>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td><strong>in</strong></td>
<td>I</td>
<td>I</td>
</tr>
</tbody>
</table>

Supersense label only at beginning of lexical segment.
AMALGrAM

• Tagger trained on STREUSLE: jointly predicts MWEs and supersenses
  ‣ $|\text{tagset}| = 146$
  ‣ Same structured prediction setup as Schneider et al. (TACL 2014): first-order structured perceptron

• Evaluation: separate scores for
  ‣ MWE identification
  ‣ supersense tagging (first tag of each lexical segment)
AMALGrAM

• This tagger allows us to measure:
  ‣ the impact of joint tagging on MWE performance
  ‣ the value of word clusters, new features
  ‣ the tagger’s resilience to ambiguity (see the paper)

• Baseline for future supersense tagging studies in the reviews domain
Outline

✓ Background
  ‣ multiword expressions
  ‣ supersenses
✓ Dataset
✓ Joint model
● Results
Does joint tagging hurt MWE identification?

- MWE-only baseline (8 tags): 73 56 63
- Simplest joint model (146 tags): 68 56 61
- ...so it hurts a bit in precision, but not drastically
AMALGrAM: New features

• **aux verb feature**: `verb (adverb)? verb`

• **WordNet features** adapted from (Ciaramita & Altun, 2006). E.g.:
  
  ▸ has-supersense (in any matching synset)
  
  ▸ supersense of 1st synset of longest lemma match
  
  ▸ (if a common noun, verb, or adjective): supersense of 1st synset matching the following noun
Impact of new features on supersense labeling

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>$F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simplest joint model (146 tags)</td>
<td>65</td>
<td>67</td>
<td>66</td>
</tr>
<tr>
<td>+ clusters</td>
<td>66</td>
<td>68</td>
<td>67</td>
</tr>
<tr>
<td>+ new features</td>
<td>69</td>
<td>72</td>
<td>71</td>
</tr>
</tbody>
</table>
Does joint tagging hurt MWE identification?

<table>
<thead>
<tr>
<th>Model Description</th>
<th>P</th>
<th>R</th>
<th>F₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>MWE-only baseline (8 tags)</td>
<td>73</td>
<td>56</td>
<td>63</td>
</tr>
<tr>
<td>Simplest joint model (146 tags)</td>
<td>68</td>
<td>56</td>
<td>61</td>
</tr>
<tr>
<td>+ clusters</td>
<td>69</td>
<td>57</td>
<td>62</td>
</tr>
<tr>
<td>+ new features</td>
<td>71</td>
<td>56</td>
<td>63</td>
</tr>
</tbody>
</table>
Conclusion

- Corpus of English web reviews annotated for MWEs + supersenses (STREUSLE)
- Tagger for this corpus attains 63% $F_1$ for MWEs and 71% $F_1$ for supersenses (with gold POS)
Possible Extensions

• More genres & languages. Already have:
  ‣ supersenses in English Twitter (Johannsen et al., 2014), Arabic Wikipedia (Schneider et al., 2012), Italian (Dei Rossi et al., 2013), ...
  ‣ some MWEs in English Wikipedia (Vincze et al., 2011), French news (Abeillé et al., 2003), ...

• More kinds of supersenses
  ‣ adjectives (Tsvetkov et al., 2014)
  ‣ prepositions (Schneider et al., LAW 2015)

• Application to sentiment analysis, semantic parsing, machine translation, ...
Links

- Downloads: tiny.cc/streusle
- Ideas for improving on this task?
  - “DiMSUM” shared task, SemEval 2016. Subscribe to mailing list for further announcements.
Many thanks
(*Several thanks)

Thanks a million
(*Thanks a thousand)

Thanks a lot
(?Lots of thanks)