Towards a Deeper Semantic Output

Zina Pozen
University of Washington
M.Sc. thesis proposal

Aspirational Goal

- Parsing based on HPSG grammars outputs not only high-precision syntactic phrase structure trees, but also Minimum Recursion Semantics representation, capturing structural semantics of a sentence.
- Augmenting it with the lexical semantic information could produce a truly rich semantic representation tantalizingly close to “natural language understanding”.

However,

The HPSG grammars lexicons are primarily designed for syntax modeling and as such are underspecified with respect to word senses, as long as the lexical items exhibit the same surface-syntactic properties.

ERG-WN noun sense mapping example

- **Only one ERG entry:**

plant_n1 := n_-_c_le & [ORTH < "plant" >, SYNSEM [LKEYS.KEYREL.PRED "_plant_n_1_rel", PHON.ONSET con]].

… for 4 in WordNet:

- <noun.artifact>S: (n) **plant#1**, works#1, industrial plant#1 (buildings for carrying on industrial labor)
- <noun.Tops>S: (n) **plant#2**, flora#2, plant life#1 ((botany) a living organism lacking the power of locomotion)
- <noun.person>S: (n) **plant#3** (an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience)
- <noun.cognition>S: (n) **plant#4** (something planted secretly for discovery by another)

Polysemy necessary for syntax modeling is present in the ERG lexicon:

fall_n1 := n_-_m-ssn_le & [ORTH < "fall" >, SYNSEM [LKEYS.KEYREL.CARG "fall", PHON.ONSET con]].

fall_n2 := n_-_m-ssn-spr_le & [ORTH < "fall" >, SYNSEM [LKEYS.KEYREL.CARG "fall", PHON.ONSET con]].

fall_n3 := n_np_m-ssn_le & [ORTH < "fall" >, SYNSEM [LKEYS.KEYREL.CARG "fall", PHON.ONSET con]].

fall_n4 := n_-_c_le & [ORTH < "fall" >, SYNSEM [LKEYS.KEYREL.PRED "_fall_n_1_rel", PHON.ONSET con]].

…but 12 senses in WordNet:

- <noun.time>S: (n) **fall#1**, autumn#1 (the season when the leaves fall from the trees)
- <noun.act>S: (n) spill#4, tumble#2, **fall#2** (a sudden drop from an upright position)
- <noun.event>S: (n) **Fall#3** (the lapse of mankind into sinfulness because of the sin of Adam and Eve)
- <noun.object>S: (n) descent#5, declivity#1, **fall#4**, decline#4, declination#2,declension#3, downslope#1 (a downward slope or bend)
- <noun.act>S: (n) **fall#5** (a lapse into sin; a loss of innocence or of chastity)
- <noun.event>S: (n) **fall#6**, downfall#3 (a sudden decline in strength or number or importance)
- <noun.event>S: (n) **fall#7** (a movement downward)
- <noun.act>S: (n) capitulation#3, **fall#8**, surrender#4 (the act of surrendering (usually under agreed conditions))
- <noun.time>S: (n) twilight#1, dusk#1, gloaming#1, gloam#1, nightfall#1, evenfall#1,**fall#9**, crepuscule#1, crepuscle#1 (the time of day immediately following sunset)
- <noun.event>S: (n) **fall#10**, pin#2 (when a wrestler's shoulders are forced to the mat)
- <noun.event>S: (n) drop#6, **fall#11** (a free and rapid descent by the force of gravity)
- <noun.attribute>S: (n) drop#3, dip#6, **fall#12**, free fall#2 (a sudden sharp decrease in some quantity)

Verbal polysemy: “take” in VerbNet

adopt-93.xml: <MEMBER name="take_over" wn="take_over%2:40:07 take_over%2:41:01" grouping=""/>
adopt-93.xml: <MEMBER name="take_on" wn="take_on%2:30:00 take_on%2:41:01" grouping=""/>

appear-48.1.1.xml: <MEMBER name="take shape" wn="" grouping=""/>
bring-11.3.xml: <MEMBER name="take" wn="take%2:38:09 take%2:38:10 take%2:42:10" grouping="take.04"/>

characterize-29.2.xml: <MEMBER name="take" wn="take%2:31:07 take%2:31:01 take%2:40:05" grouping="take.05 take.07 take.08"/>
confront-98.xml: <MEMBER name="take_on" wn="take_on%2:41:00" grouping=""/>
convert-26.6.2.xml: <MEMBER name="take" wn="take%2:41:13" grouping="take.06"/>
cost-54.2.xml: <MEMBER name="take" wn="take%2:40:06" grouping="take.03"/>
fit-54.3.xml: <MEMBER name="take" wn="take%2:42:15" grouping="take.10"/>
hire-13.5.3.xml: <MEMBER name="take" wn="take%2:40:03" grouping="take.05"/>
occurrence-48.3.xml: <MEMBER name="take place" wn="take_place%2:30:00" grouping=""/>
performance-26.7.xml: <MEMBER name="take" wn="take%2:32:02" grouping="take.01 take.02"/>
rely-70.xml: <MEMBER name="take_a_chance" wn="take_a_chance%2:41:00" grouping=""/>
require-103.xml: <MEMBER name="take" wn="take%2:42:00" grouping="take.07"/>
steal-10.5.xml: <MEMBER name="take" wn="?take%2:38:09 take%2:40:01" grouping="take.04 take.09"/>

“take” in WordNet – 42 senses across 12 lexical files

1	**<verb.body>**
4	**<verb.change>**
5	**verb.cognition>**
3	**<verb.communication>**
2	**<verb.competition>**
1	**<verb.consumption>**
2	**<verb.contact>**
6	**<verb.motion>**
1	**<verb.perception>**
10	**<verb.possession>**
3	**<verb.social>**
4	**<verb.stative>**

“take” in the ERG – phrasal modeling, only two predicates for non-phrasal

1 KEYREL.PRED "_take_v_aback_rel"],
1 KEYREL.PRED "_take_v_along_rel"],
1 KEYREL.PRED "_take_v_apart_rel"],
1 KEYREL.PRED "_take_v_around_rel"],
1 KEYREL.PRED "_take_v_away_rel"],
1 KEYREL.PRED "_take_v_back_rel"],
1 KEYREL.PRED "_take_v_down_rel"],
1 KEYREL.PRED "_take_v_home_rel"],
1 KEYREL.PRED "_take_v_i_rel"],
1 KEYREL.PRED "_take_v_in_rel"],
1 KEYREL.PRED "_take_v_into_rel"],
1 KEYREL.PRED "_take_v_off_rel"],
1 KEYREL.PRED "_take_v_on_rel"],
1 KEYREL.PRED "_take_v_out+of_rel"],
1 KEYREL.PRED "_take_v_to_rel"],
1 KEYREL.PRED "_take_v_up_rel"],
1 KEYREL.PRED "_take_v_x-off_rel"],
1 SYNSEM [LKEYS.KEYREL.PRED "_take_n_1_rel",
1 SYNSEM [LKEYS.KEYREL.PRED "_take_v_of-i_rel",
2 KEYREL.PRED "_take_v_of-i_rel"],
2 KEYREL.PRED "_take_v_out_rel"],
2 KEYREL.PRED "_take_v_over_rel"],
2 SYNSEM [LKEYS.KEYREL.PRED "_double-take_n_1_rel",
3 SYNSEM [LKEYS.KEYREL.PRED "_take_v_1_rel",
3 SYNSEM [LKEYS.KEYREL.PRED "_take_v_2_rel",

Some ERG lexicon experiments

How much polysemy is there in the current ERG
(~ how much polysemy is syntactically expressed in the modern American English ?)

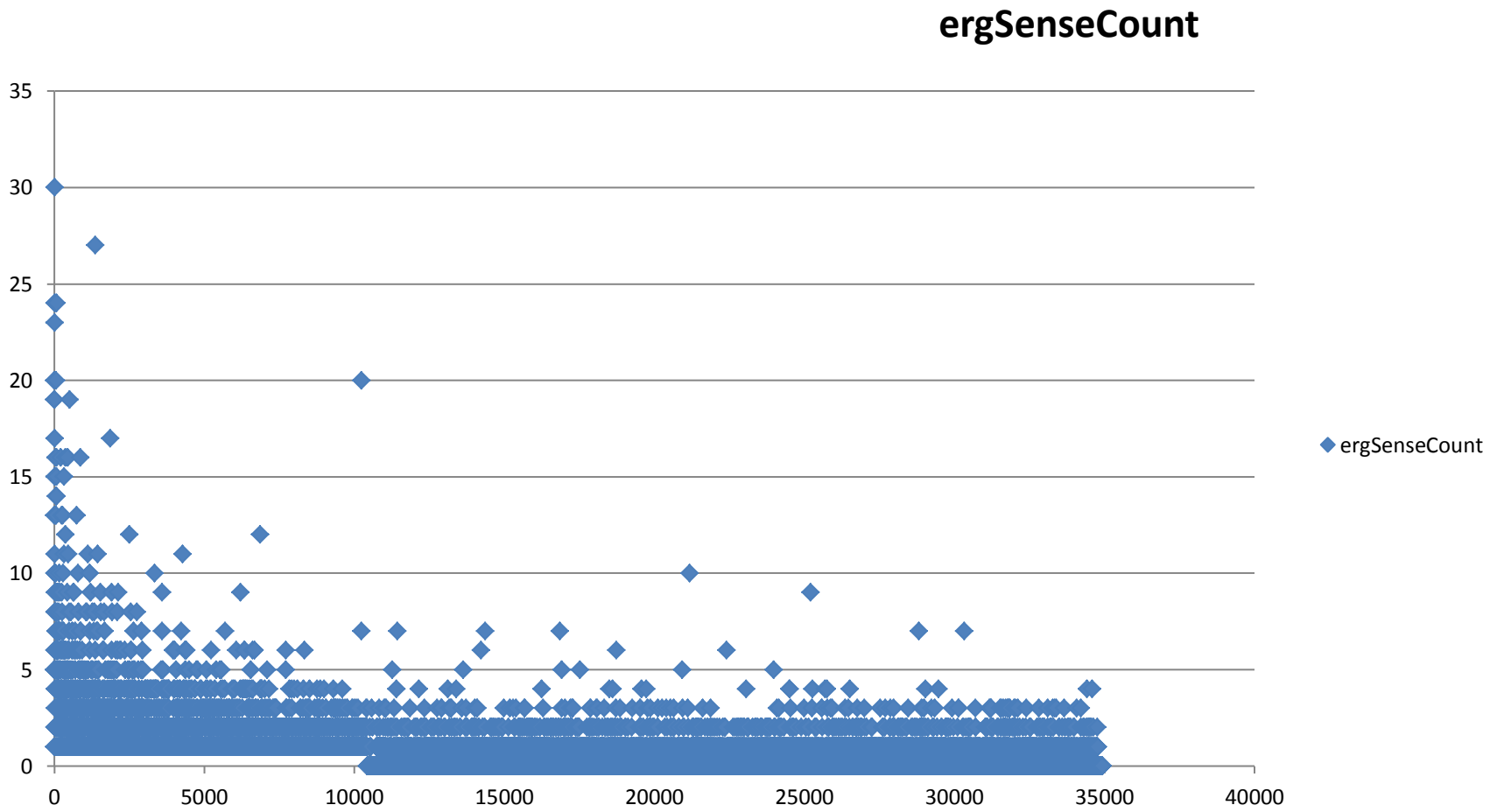
Nouns with 2+ senses – **720/18157 (3.965 %)**

Verbs with 2+ senses – **1785/8229 (21.692%)**

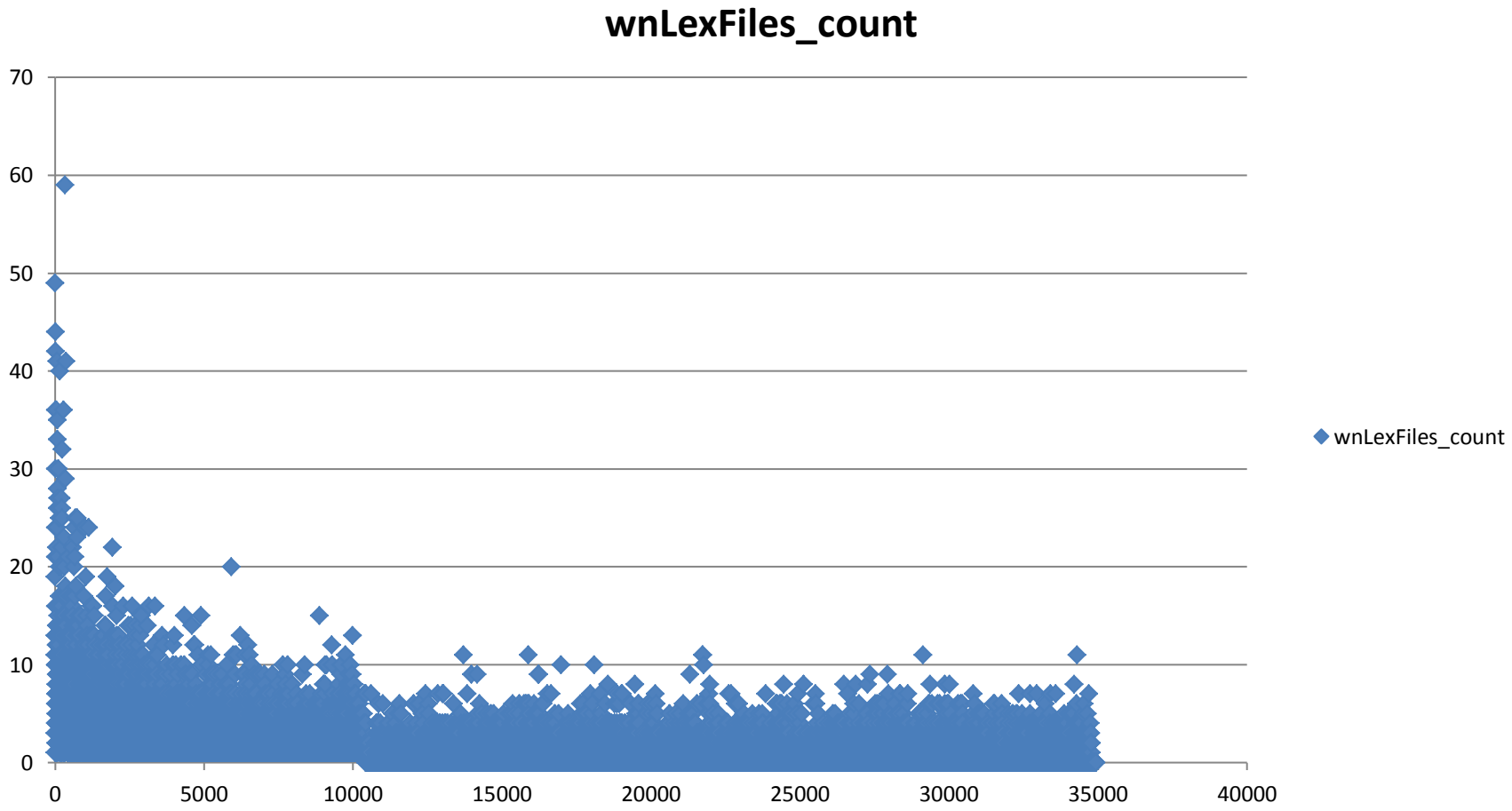
Adjectives – **610/5603 (10.887%)**

Adverbs - **205/2054 (9.980%)**

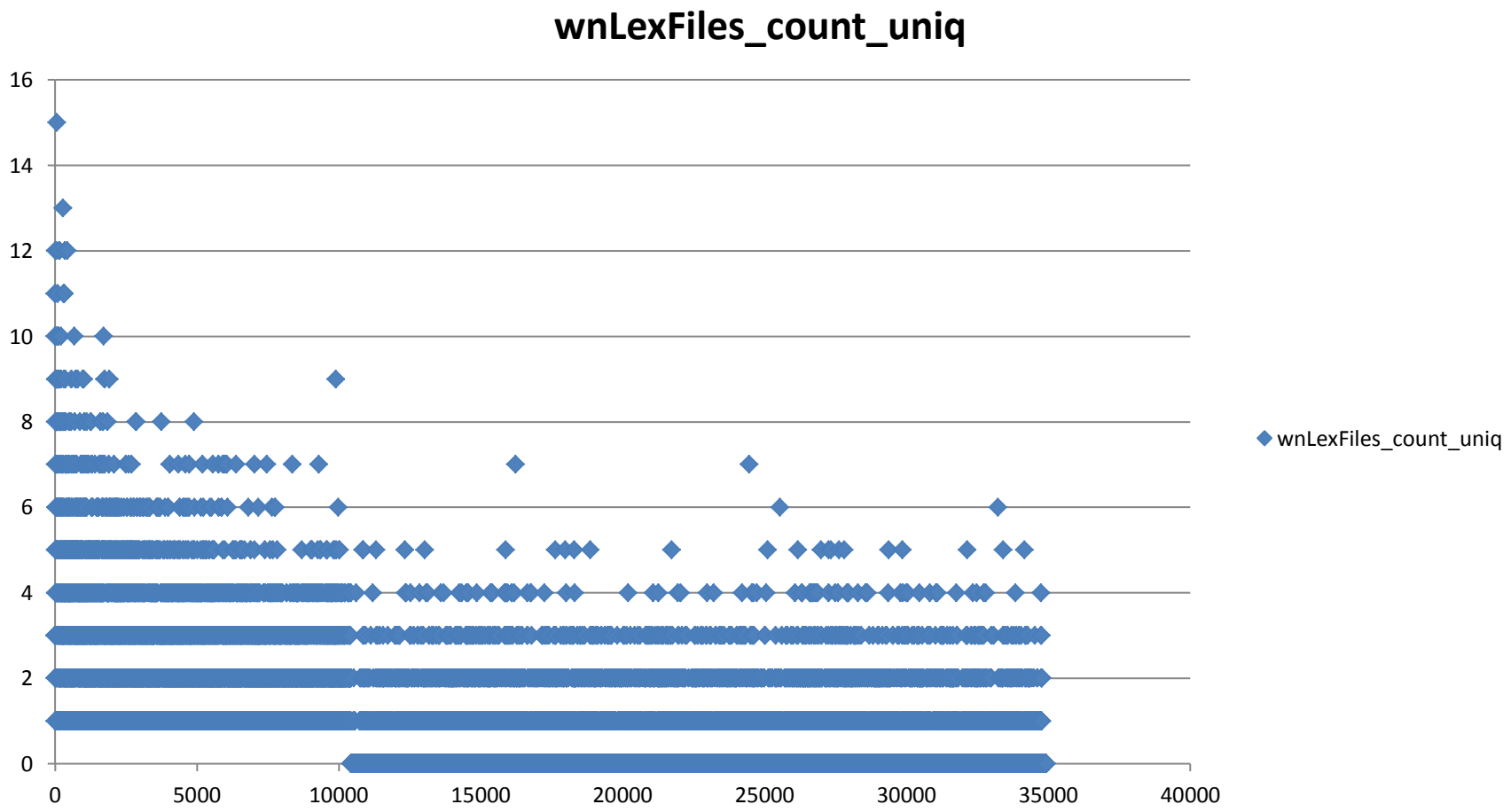
ERG sense count – sorted by WN frequency



WN sense count for ERG lexicon – sorted by WN frequency



WN unique coarse (lex file) sense count, sorted by frequency



How can we disambiguate ERG output?

- Marking the senses in the lexicon will explode parse numbers – unless we include sense selectional restrictions to predications.
- The restrictions would probably have to be probability rather than unification-based.
- Interested in learning word-to-word, word-to-class, class-to-class selectional preferences over EPs
- Classes could be WN lex files or LCSs, Levin classes for verbs, possibly others…

Potential Resources

Redwoods 7th growth includes Semcor 3.0 - opportunity for training and evaluation (but need to figure out item alignment).

May need more data, esp. for word-to-word models – may harvest unambiguous EPs from untagged Redwoods corpora

Perhaps take advantage of VerbNet frames when predicates are unambiguous

Recent related work

Fujita et al., 2010 -

Fujita, S., Bond, F., Oepen, S., & Tanaka, T. (January 01, 2010). Exploiting Semantic Information for HPSG Parse Selection. *Research on Language and Computation, 8,* 1, 1-22.