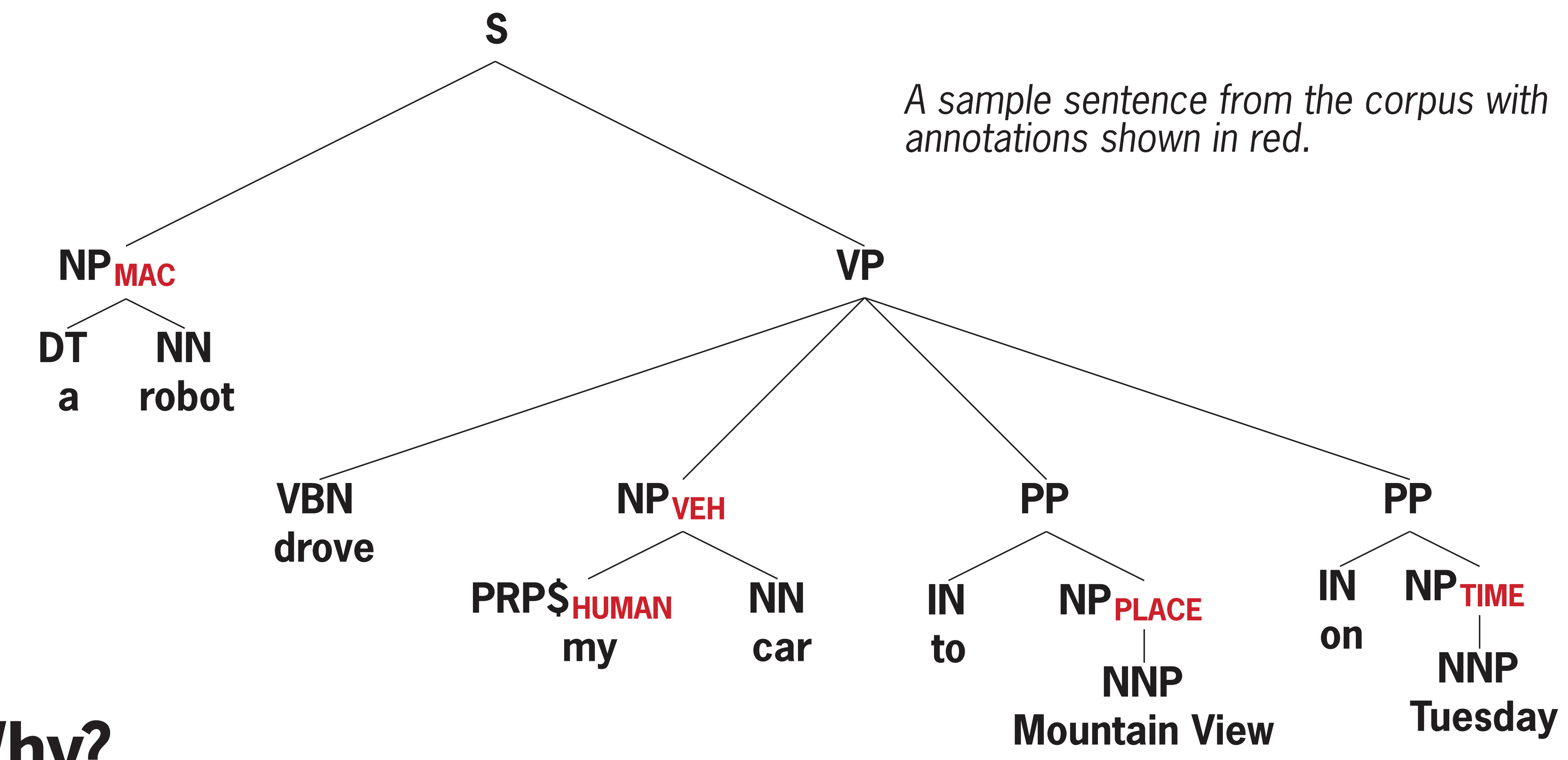
Automatic animacy classification

Samuel Bowman[1] and Harshit Chopra[2]

[1]Department of Linguistics, Stanford University [2]Department of Computer Science, Stanford University

What?

We introduce the automatic annotation of noun phrases in parsed sentences with tags from a fine-grained semantic animacy hierarchy. These tags reflect an important lexical semantic property, and show promise as features for a number of NLP tasks.



A sample sentence from the corpus with annotations shown in red.

Why?

- The classes capture the degree to which the entity described by an NP is capable of **human-like volition**.
- Major predictor of verbal argument selection; triggers a range of morphological and syntactic phenomena across languages (Levin and Rappaport Hovav, 2005).
- Annotating a corpus with this information can facilitate:
  - **Natural language generation**
  - **Statistical language modeling**
  - **Parse selection**
  - **Machine translation**
  - **Corpus lexical semantics**

  (Zaenen et al., 2004; Øvrelid and Nivre, 2007)
- All existing classification work classifies only the basic ANIMATE/INANIMATE contrast (Ji and Lin, 2009; Øvrelid, 2005; Orasan and Evans, 2001).
- All existing work on animacy in English uses outside lexical resources.

Classes and Corpus

- **Zaenen et al. (2004)**'s annotation scheme and corpus:
  - Ten classes: HUMAN, ORG (organizations), ANIMAL, MAC (automata), VEH (vehicles), PLACE, TIME, CONCRETE (other physical objects), NONCONC (abstract entities), and MIX (heterogeneous groups).
  - An annotated subset of the hand-parsed NXT **Switchboard corpus of conversational American English** (Calhoun et al., 2010).
  - About 110,000 sentences with about 300,000 NPs.
- **Data division**: training (80%), development (10%), test (10%)

  Note: Some feature selection was inadvertently done before this split was finalized. All relevant experiments have been repeated on the current split.

Model and Features

- **Maximum entropy classifier** (Berger et al., 1996) with three feature bundles:
  - **Bag of words features** capture every word in the NP:

    -HASWD-(*POS-tag*-)*word*

    "the mayor"

    → {HASWD-*DT*-*the*, HASWD-*the*, HASWD-*NN*-*mayor*, HASWD-*mayor*}
  - **Internal syntactic features** reflect that the head of an NP typically carries the bulk of the information on animacy. Adding orthographic shape helps with unseen words.

    -HEAD-*tag*-*word*

    -HEADSHAPE-*tag*-*shape*

    "The Panama hat I gave the mayor"

    → {HEAD-NN-hat, HEADSHAPE-NN-L}
  - **External syntactic features** reflect that verbs and prepositions tend to restrict the classes of their arguments:

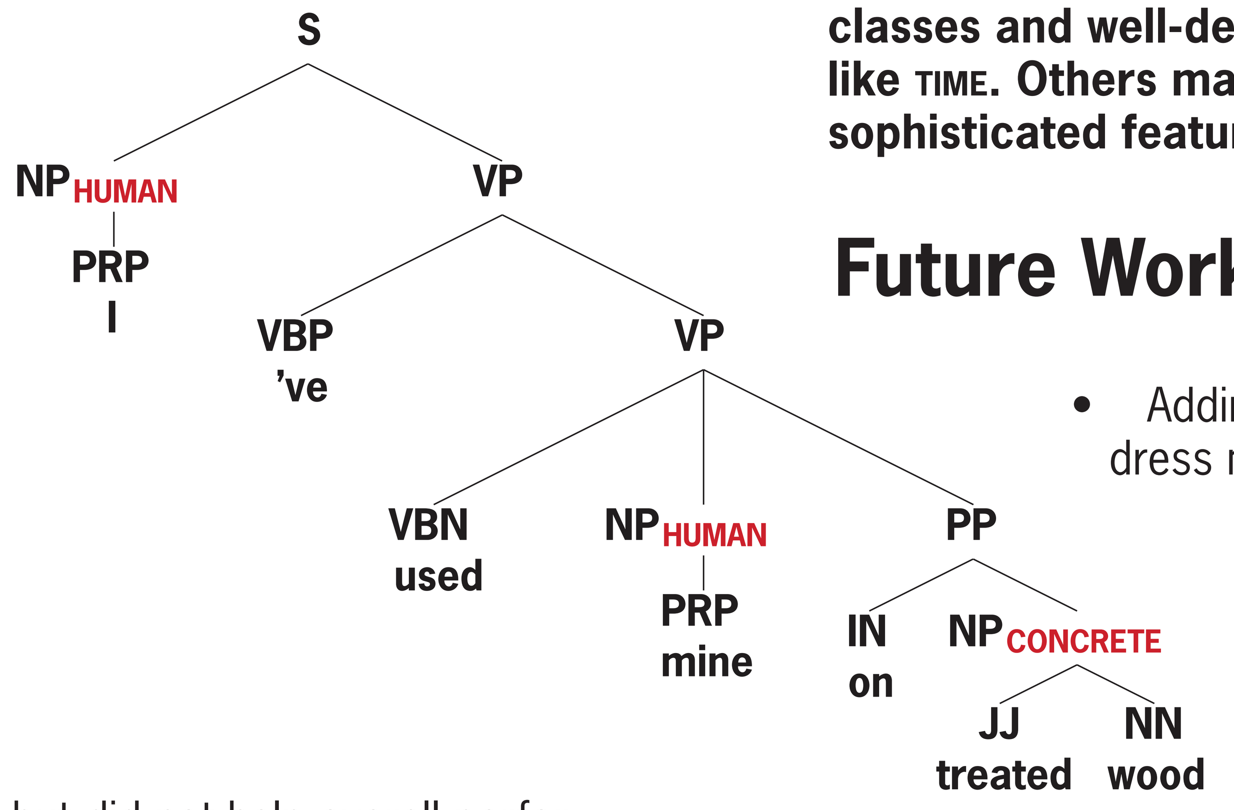
    -SUBJ(-OF-*verb*)

    -DOBJ(-OF- *verb*)

    -PCOMP(-OF-*prep*)(-WITH-*verb*)

    "I called [the mayor]$_{NP}$"

    → {DOBJ, DOBJ-OF-called}



- Features which introduced limited **dependencies between classes** helped with MIX NPs, but did not help overall performance, and were scrapped.

Results and Discussion

- Our **baseline** always chooses the most frequent class, NONCONC.
- **Binary ANIMATE/INANIMATE classification**: 93.50% accuracy. Baseline labeling each NP ANIMATE: 53.79%.
- **Automatically parsing** the corpus with the Stanford parser (Klein and Manning, 2002) generated correct NPs with Pr. 88.63% / Rec. 73.51%. For these NPs: 85.43% accuracy.
- Many **errors** from pronouns whose referents are not specified within the sentence:
  - In **the tree below**, for example, the model wrongly, but plausibly, classified "mine" as NONCONC.
- Subtle distinction between plural HUMAN (an incidental group) and ORG (a group with voice or purpose).
- **High accuracy on common classes and well-defined classes like TIME. Others may need more sophisticated features.**

Class	Count	Precision	Recall
VEH	534	88.56	39.14
TIME	1,101	88.24	80.38
NONCONC	12,173	83.39	93.32
MAC	79	63.33	24.05
PLACE	754	64.89	63.00
ORG	1,208	58.26	27.73
MIX	29	7.14	3.45
CONCRETE	1402	58.82	37.58
ANIMAL	137	69.44	18.25
HUMAN	11,320	91.19	93.30
Overall	28,737	Accuracy: 84.90	

Counts and performance for each class

Only these features:	Accuracy (%)
Bag of words	83.04
Internal Syntactic	75.85
External Syntactic	50.35
All but these features:	—
Bag of words	77.02
Internal syntactic	83.36
External syntactic	84.58
Most frequent class	42.36
Full model	**84.90**

Performance for each feature bundle alone, and with each feature bundle removed

Future Work

- Adding **coreference resolution** between sentences would address many errors without requiring outside data sources.
- Features from **WordNet** (Fellbaum, 2010) and **FrameNet** (Baker et al., 1998):
  - Synonyms and hypernyms would help with unknown words (Orasan And Evans 2001).
  - **Semantic role labels** would help to capture verbal animacy restrictions. Might rescue the relatively ineffective external syntactic features.

Acknowledgements

We are indebted to Marie-Catherine de Marneffe and Jason Grafmiller, who first suggested we model this corpus, and to Chris Manning, Gabor Angeli, and our reviewers for valuable advice.

Questions: sbowman@stanford.edu

References

C.F. Baker, C.J. Fillmore, and J.B. Lowe. 1998. The Berkeley Framenet Project. Proc. of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics.

A.L. Berger, V.J Della Pietra, and S.A. Della Pietra. 1996. A maximum entropy approach to natural language processing. Computational Linguistics, 22(1).

S. Calhoun, J. Carletta, J.M. Brenier, N. Mayo, D. Jurafsky, M. Steedman, and D. Beaver. 2010. The NXT-format Switchboard Corpus. Language resources and evaluation, 44(4).

C. Fellbaum. 2010. Wordnet. In Theory and Applications of Ontology: Computer Applications. Springer.

H. Ji and D. Lin. 2009. Gender and animacy knowledge discovery from web-scale N-grams for unsupervised person mention detection. Proc. of the 23rd Pacific Asia Conference on Language, Information and Computation.

D. Klein and C.D. Manning. 2002. Fast exact inference with a factored model for natural language parsing. Advances in neural information processing systems, 15(2002).

B. Levin and M. Rappaport Hovav. 2005. Argument Realization. Cambridge.

C. Orasan and R. Evans. 2001. Learning to identify animate references. Proc. of the Workshop on Computational Natural Language Learning, 7.

L. Øvrelid and J. Nivre. 2007. When word order and part-of-speech tags are not enough—Swedish dependency parsing with rich linguistic features. In Proc. of the International Conference on Recent Advances in Natural Language Processing.

L. Øvrelid. 2005. Animacy classification based on morphosyntactic corpus frequencies: some experiments with Norwegian nouns. In Proc. of the Workshop on Exploring Syntactically Annotated Corpora.

A. Zaenen, J. Carletta, G. Garretson, J. Bresnan, A. Koontz-Garboden, T. Nikitina, M.C. O'Connor, and T. Wasow. 2004. Animacy encoding in English: Why and how. In Proc. of the Association for Computational Linguistics Workshop on Discourse Annotation.