Designing Agreement Features for Realization Ranking

Rajakrishnan Rajkumar and Michael White
Department of Linguistics, The Ohio State University

Summary

- We show that incorporating linguistically motivated features to ensure correct animacy, number agreement and punctuation balancing in an averaged perceptron ranking model for CCG realization helps improve a state-of-the-art baseline even further
- Traditionally, these features have been modelled using hard constraints in the grammar
- Given the graded nature of grammaticality judgements in the case of animacy and the heterogenous nature of agreement, we argue a case for the use of a statistical model to rank competing preferences
- Compared to writing grammar rules, our method is more robust and allows incorporating information from diverse sources in realization
- Our efforts result in fewer agreement errors and yield a state-of-the-art BLEU score of **0.8446** on Section 23 of the CCGbank

The Graded Nature of Animacy Agreement

- In English, an inanimate noun can be modified by a relative clause introduced by *that* or *which* while an animate noun combines with *who(m)*
- With some nouns though — such as *team*, *group*, *squad*, etc. — animacy status is uncertain, and these can be found with all the three relative pronouns (*who, which* and *that*)
- Google counts suggest that all three choices are almost equally acceptable
- OpenCCG takes as input LFs underspecified for relative pronouns, thus leaving the ranking model to enforce the correct choice

The Heterogeneous Nature of Number Agreement

- Subject-verb agreement has often been considered to be a syntactic phenomenon and grammar implementations generally use syntactic features to enforce agreement constraints (e.g. Velldal and Oepen, 2005)
- However a closer look at our data and a survey of the theoretical linguistics literature (Kathol, 1999; Pollard and Sag, 1994) points toward a more heterogeneous conception of English agreement:
  - (1) " I think it will shake confidence one more time , and a lot of this business is based on client confidence . " (WSJ1866.10)
  - (2) It 's interesting to find that a lot of the expensive wines are n't always walking out the door . (WSJ0071.53)
- Kathol (1999) proposes an explanation where agreement is determined by the semantic properties of the noun rather than by its morphological properties
- However, the semantic view of agreement is not completely convincing due to counterexamples like the following discussed in the literature (reported in Kim (2004)):
  - (3) Suppose you meet someone and they are totally full of themselves
  - (4) Those scissors are missing.
- We investigate here the extent to which a machine learning–based approach is a simpler, practical alternative for acquiring the relevant generalizations from the data by combining information from various information sources

CCG Surface Realization

Logical Forms and Hypertagging

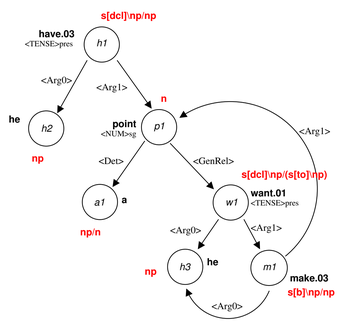
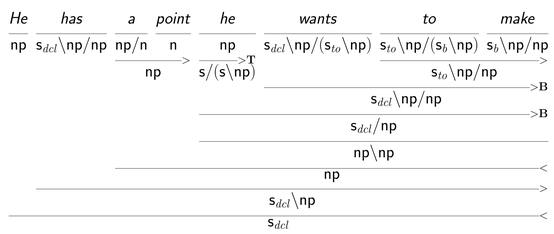


Chart Realization



- Logical forms are underspecified for relative pronouns and case marking prepositions
- Realizations are ranked using three language models and a perceptron model

Presenting Balanced Punctuation

The task of ensuring the correct realization of commas introducing noun phrase appositives:

(5) John, CEO of ABC, loves Mary.
(6) * John, CEO of ABC loves Mary.
(7) Mary loves John, CEO of ABC.
(8) * Mary loves John, CEO of ABC,.
(9) Mary loves John, CEO of ABC, madly.
(10) * Mary loves John, CEO of ABC madly.

- Using syntactic features to block unwanted derivations (Briscoe, 1994) does not work for CCG
- We modified the result categories of unbalanced appositive commas and dashes to include a feature marking unbalanced punctuation
- During feature extraction, derivations were examined to detect categories such as $\mathsf{np}_{unbal=comma}$, and checked to make sure this NP is followed by another punctuation mark in the string such as a full stop

OpenCCG Realization Results

Legend

Model	Description
full-model	All the feats from models below
agr-punct	Baseline Feats + Punct + Num-Agr
wh-punct	Baseline Feats + Punct + Animacy-Agr
baseline-punct	Baseline Feats + Punct
baseline	Log prob + n-gram +Syntactic features

Section 00 Error Analysis

Model	#Punct-Errs	%Agr-Errs	%WH-Errs
baseline	39	11.05	22.44
baseline-punct	0	10.79	20.77
wh-punct	11	10.87	13.53
agr-punct	8	4.0	21.84
full-model	10	4.31	15.53

Section	Model	%Exact	%Compl.	BLEU
00	baseline	38.18	82.47	0.8341
	baseline-punct	37.97	82.47	0.8340
	wh-punct	38.93	82.53	0.8360
	full-model	40.47	82.53	0.8403
	agr-punct	40.84	82.53	**0.8414**
23	baseline	38.98	83.39	0.8442
	full-model	40.09	83.35	**0.8446**

Feature Design

Features in the Baseline Perceptron Ranking Model (White and Rajkumar 2009)

The baseline perceptron ranking model has the following features:

- The log probability of the candidate realization's word sequence according to our linearly interpolated language models as a single feature in the perceptron model (Velldal & Oepen 2005)
- Since our language model linearly interpolates three component models, we also include the log prob from each component language model as a feature, so that the combination of these components can be optimized
- Syntactic features in our model are from Clark & Curran's (2007) normal form model
- Discriminative n-gram features (Roark et al., 2004), which count the occurrences of each n-gram that is scored by our factored language model

However an error analysis revealed that the above features were inadequate to avoid animacy, number and balanced punctuation errors.

Designing New Features

We designed the new features shown in the table alongside:

- During feature extraction, subjects of the verbs tagged VBZ and VBP are identified using the PTB NP-SBJ function tag annotation projected on to the appropriate arguments of lexical categories of verbs
- The first class of number agreement features encode all possible combinations of subject-verb word forms and parts of speech tags. This is to deal with simple cases of subject verb agreement.
- In the case of NPs involving *of*-complements like *a lot of ...* (Examples 1 and 2), feature classes 2 and 3 were extracted (class 1 was excluded)
- Writing grammar rules to get these facts right proved to be surprisingly difficult and required a list of measure nouns and partitive quantifiers
- Feature extraction is also sensitive to both conjunct/disjunct VPs and subject NPs

Feature	Example
Animacy features	
Noun Stem + Wh-pronoun	researcher + who
Noun Class + Wh-pronoun	PER_DESC + who
Number features	
Noun + Verb	people + are
NounPOS + Verb	NNS + are
Noun + VerbPOS	people + VBP
NounPOS + VerbPOS	NNS + VBP
Noun_of + Verb	lot_of + are
Noun_of + VerbPOS	lot_of + VBP
NounPOS_of + Verb	NN_of + are
NounPOS_of + VerbPOS	NN_of + VBP
Noun_of + *of*-complementPOS + VerbPOS	lot_of + NN + VBZ
NounPOS_of + *of*-complementPOS + VerbPOS	NN_of + NN + VBZ
Noun_of + *of*-complementPOS + Verb	lot_of + NN + is
NounPOS_of + *of*-complementPOS + Verb	NN_of + NN + is
Punctuation feature	
Balanced Punctuation Indicator	$unbalPunct=1

Realizer Evaluation

A Section 02–21 lexico-grammar extracted from the CCGbank (with NE collapsing) is used to realize Section 00 and Section 23 LFs. The ungrammatical examples below produced by the baseline ranking model are realized correctly by the model with all the features.

(11) Neither Lorillard nor the researchers who/*that studied the workers were aware of any research on smokers of the Kent cigarettes ...
(12) While many of the risks were/*was anticipated when minneapolis-based Cray Research first announced the spinoff ...
(13) Giant Group is led by three Rally 's directors , Burt Sugarman , James M. Trotter III and William E. Trotter II that last month indicated that they hold/*holds a 42.5 % stake in Rally 's and plan/*plans to seek a majority of seats on ...
(14) ... the ban wo n't stop tissue-transplant privately funded research or federally funded fetal-tissue **research** that does n't/*do n't involve transplants .

Acknowledgements

This work was supported in part by NSF grant IIS-0812297 and by an allocation of computing time from the Ohio Supercomputer Center. Our thanks also to Robert Levine and the anonymous reviewers for helpful comments and discussion.