Automatic Grammaticality Judgements

Joachim Wagner

NCLT Seminar Series 21st Mar 2007

Talk Outline

- Motivation
- Approaches
  - XLE default grammar
  - PTB-trained PCFGs
  - POS n-grams
  - Parse probability
- Test corpus
- Results, Conclusions, Future Work

Why Judge the Grammaticality?

- Automatic essay grading
- Trigger deep error analysis
  - Increase speed
  - Reduce overflagging
- Most approaches easily extend to
  - Locating errors
  - Classifying errors

Why this Evaluation?

- No agreed standard
- Differences in
  - What is evaluated
  - Corpora
  - Error density
  - Error types

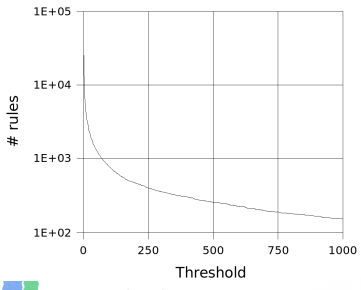
Approach 1: XLE Grammar

- Default English LFG installed on zag
- Optimality Theory
  - Ranking of constraints
  - Dispreferred rules
- XLE parser
  - “starred” sentences
  - Number of optimal and “unoptimal” parses

Approach 2: PTB-trained PCFG

- Trained on WSJ section 02-23 (Charniak 1996, Cahill 2004)
- Parses almost any input
- Prune grammar rules
  - Reduce robustness
  - Frequency threshold

Approach 2: PTB-trained PCFG



Approach 3: n-grams

- Flag unlikely or rare sequences
  - POS (different tagsets)
  - Tokens
  - Raw frequency vs. mutual information
- Most publications are in the area of context-sensitive spelling correction
  - Real word errors
  - Resulting sentence can be grammatical

Approach 4: Parse probability

- PTB-trained, robust, probabilistic parser
  - Probability of expanding start symbol to given tree
- Reference probability
  - Assuming grammatical input
  - Estimate using low-level features
  - Machine learning algorithms: k-NN
  - Training data: grammatical data only

Test Corpus

- BNC w\o speech, poems, captions, lists
- Randomly selected 10 subsets
  - Cross-validation
  - Sentence length 10 – 19
- Automatically inserted errors
  - Agreement errors
  - Context-sensitive spelling errors
  - Extra word errors
  - Missing word errors
- 1:1 grammatical and ungrammatical test data

Test Corpus - Example

- Missing Word Error

She didn't **want** to face him

↓

She didn't to face him

Test Corpus – Example 2

- Context-sensitive spelling error

I love **them** both

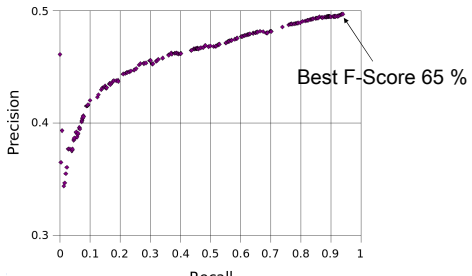
↓

I love **then** both

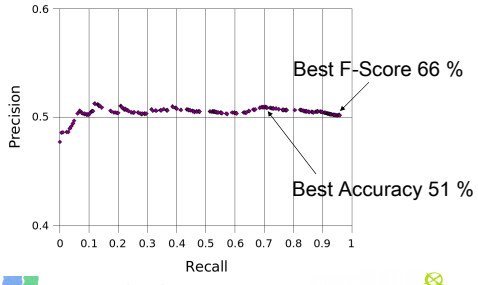
Results 1: XLE parser

	Precision	Recall	F-Score
Agreement	71%	60%	65%
CS Spelling	69%	53%	60%
Extra Word	70%	57%	63%
Missing Word	63%	41%	50%

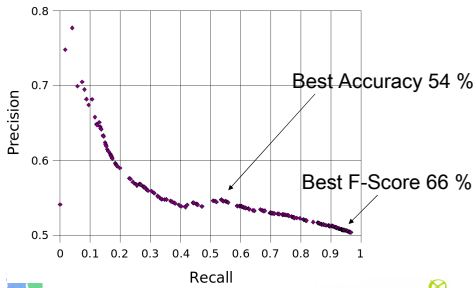
Results 2: Pruned PCFG and Agreement Errors



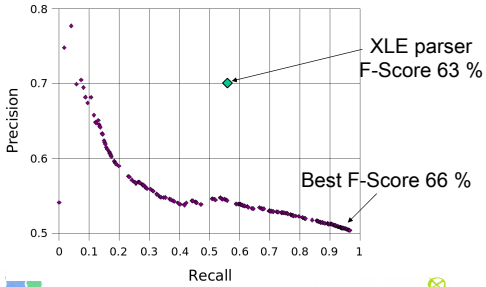
Results 2: Pruned PCFG and Context-Sensitive Spelling Errors



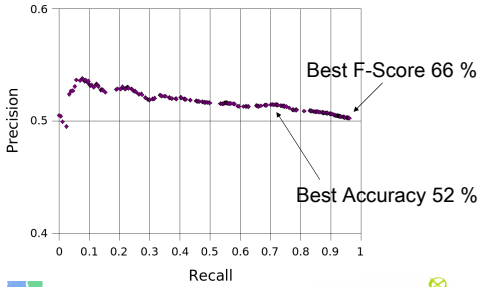
Results 2: Pruned PCFG and Extra Word Errors



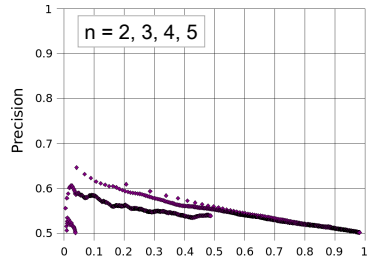
Results 2: Pruned PCFG and Extra Word Errors



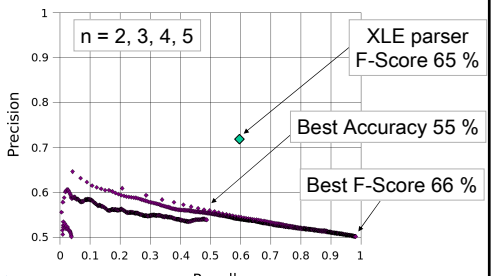
Results 2: Pruned PCFG and Missing Word Errors



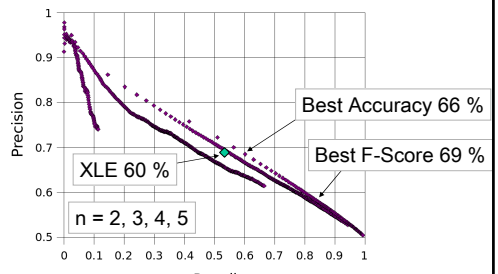
Results 3: POS n-grams and Agreement Errors



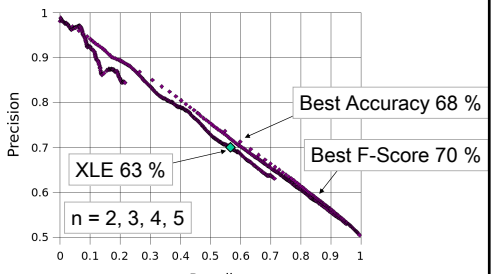
Results 3: POS n-grams and Agreement Errors



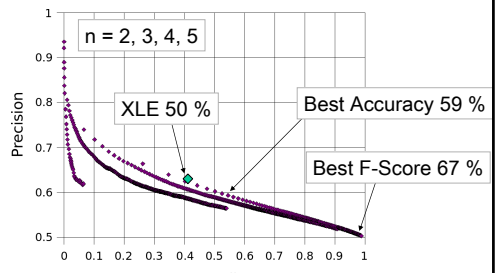
Results 3: POS n-grams and Context-Sensitive Spelling Errors



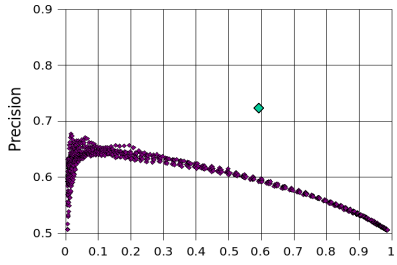
Results 3: POS n-grams and Extra Word Errors



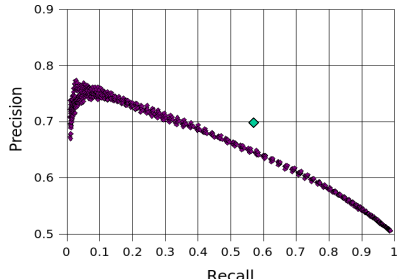
Results 3: POS n-grams and Missing Word Errors



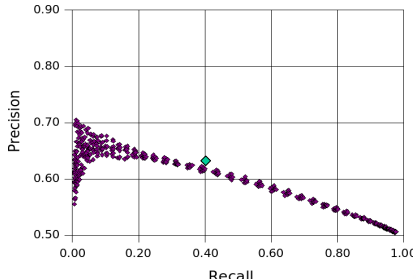
Results 4: Parse Probability and Agreement Errors



Results 4: Parse Probability and Extra Word Errors



Results 4: Parse Probability and Missing Word Errors



Conclusions

- F-Score not useful for parameterised approaches (baseline = 2/3)
- XLE surprisingly bad
- n-gram method slightly
  - except for agreement errors
- Own method below n-gram and XLE

Future Work

- Noisy channel model (Mays et al. 1991)
  - Input S, candidate correction S'
  - Maximise P(S') * P(S|S')
- Train classifier with features
  - Frequency of rarest n-gram
  - Highest threshold for which pruned PCFG still parses input
  - XLE: # optimal and "unoptimal" parses

References

- Atwell (1987): How to detect grammatical errors in a text without parsing it. In Proc. of the 3rd European chap. of the ACL, 38-45
- Brill and Moore (2000): An improved error model for noisy channel spelling correction. In Proc. of the 38th Ann. M. on ACL
- Cahill (2004): Parsing with Automatically Acquired, Wide-Coverage, Robust, Probabilistic LFG Approximations. Doctoral thesis, Dublin City University.
- Charniak (1996): Tree-Bank Grammars. In Proc. of the 19th National Conf. on AI, pp. 1031-1036, Menlo Park, CA
- Mays et al. (1991): Context based spelling correction. In Information Processing and Management, 23 (5), pp. 517-522
- Wagner et al. (2006): Detecting Grammatical Errors Using Probabilistic Parsing. Workshop on Interfaces of Intelligent Computer-Assisted Language Learning, Ohio State University

Thank you!

- Any questions?