



WordQ™ & SpeakQ™



Grammaticality Judgement in a Word Completion Task

Alfred Renaud, Fraser Shein, and Vivian Tsang

Word Completion/Prediction

- Provides writing support by suggesting words as a user types
- Benefits users of all ages, including those with disabilities, who struggle with writing and editing
- Moves user beyond basic mechanical aspects of writing to creative aspects of writing
- Increases user's independence, self-confidence, motivation

Syntax

- **WordQ: Our Word Completion System**
 - Word bigram model can lead to violations of syntax
- **Partial Parsing**
 - No luxury of waiting for complete sentence
- **Our Symbolic Approach**
 - Implement rule-based ATN parser using CFG
 - Find syntactic features current word must possess
 - Filter implausible word choices from frequency-ranked list our core predictor generates via bigrams

Grammaticality Judgement Experiment

- **Goal**

- Evaluate impact of our parser-in-progress
- Use human judgement on partial sentences as basis of evaluation

- **Methodology**

- Invite people to judge syntactic correctness of system output
 - Present sentence fragments with lists of predicted words
 - Ask subjects to identify unacceptable words in each list

- **Design**

- 3 systems: WordQ – syntax filter, WordQ + syntax filter, competitor product CoWriter
- 14 subjects
- 123 unique sentences, 41 per system
- 9 categories of linguistic challenges to guide truncation
- 5 predicted word choices per sentence

WordQ- Example

People with this condition h__

1. had
2. had been
3. has
4. has been
5. he

Results & Analysis

- **Subject-to-system agreement with 3 systems**
 - 68.1% WordQ- (70.9% post-filtered: applying our parser a posteriori)
 - 72.5% WordQ+
 - 73.6% Competitor (75.6% post-filtered)
- **Inter-subject agreement**
 - ~85% reliability across each system
 - Practical upper bound on subject-to-system agreement *any* system can achieve against this study's data

WordQ+ Example

You foster an environment that a__

1. a	13	93%	93%
2. all	13	93%	93%
3. an	14	100%	100%
4. as	8	57%	57%
5. at	8	57%	57%
	<i>11.2</i>	<i>80%</i>	<i>80%</i>

Subject-to-System

Inter-Subject

WordQ⁺ Example: Unfiltered

You foster an environment that a__

1. a	13	93%	93%
2. all	13	93%	93%
3. are	?	?	?
4. as	8	57%	57%
5. at	8	57%	57%

Subject-to-System

Inter-Subject

Competitor Example

Japanese farmers immediately pick the shoots which a__

1. a	13	93%	93%
2. am	1	7%	93%
3. and	1	7%	93%
4. are	14	100%	100%
5. aren't	14	100%	100%
	8.6	61%	96%

Subject-to-System

Inter-Subject

Competitor Example: Post-Filtered

Japanese farmers immediately pick the shoots which a__

1. a	13	93%	93%
2. am	13	93%	93%
3. and	13	93%	93%
4. are	14	100%	100%
5. aren't	14	100%	100%
	13.4	96%	96%

Subject-to-System

Inter-Subject

WordQ- Example, Revisited

People with this condition h__

1. had	13	93%	93%
2. had been	13	93%	93%
3. has	1	7%	93%
4. has been	1	7%	93%
5. he	5	36%	64%
	6.6	47%	87%

Subject-to-System

Inter-Subject

WordQ- Example, Revisited: Post-Filtered

People with this condition h__

1. had	13	93%	93%
2. had been	13	93%	93%
3. has	13	93%	93%
4. has been	13	93%	93%
5. he	5	36%	64%
	11.4	81%	87%

Subject-to-System

Inter-Subject

Discussion

- **Subjects evaluate grammaticality as readers**
 - May not anticipate broader syntactic structures when reading incomplete utterances from word to word
- **But the writer bears a unique syntactic perspective**
 - e.g., *either* people [with this condition] have..,
or people [with this condition **he** (described to us..)],
not both



Conclusion

- **Our Parser's Performance**
 - Better than bigram-only model (72.5% vs. 68.1%, $p < .05$)
 - Comparable to competitor (72.5% vs. 73.6%, $p = .212$)
 - But syntax alone cannot do it all
- **Future Work**
 - Revise experimental setup?