Syntactic Disambiguation for the Semantic Web

Jonathan Pool
Turing Center, University of Washington
Seattle, Washington, USA
pool@cs.washington.edu

S. M. Colowick
Utilika Foundation
Seattle, Washington, USA
smc@utilika.org

ABSTRACT

Are people willing and able to disambiguate content for the Semantic Web? We asked subjects to use two methods (paraphrasal and truth-conditional selection) to disambiguate sentences from the Web. Native speakers did better with the paraphrasal method, and non-native speakers with the truth-conditional method. Unpaid volunteers performed better than paid subjects. Subjects' average disambiguation time was about 20 seconds per sentence.

Categories and Subject Descriptors

H.1.2 User/Machine Systems – *Human factors, human information processi*ng

H.5.2 User Interfaces – *Natural language*

I.2.4 Knowledge Representation Formalisms and Methods – *Semantic networks*

I.2.6 Learning – *Knowledge acquisition*

I.7.2 Documentation Preparation – *Markup languages*

J.5 Arts and Humanities – *Linguistics*

General Terms

Economics, Experimentation, Human Factors, Languages

Keywords

Ambiguity, Annotation, Disambiguation, Distributed Human Computation, Metadata, Semantic Web

INTRODUCTION

Ambiguity and vagueness pervade the unstructured Web. The Semantic Web initiative proposes to rely on humans to create unambiguous content, metadata, and queries, but people have limited ability to recognize and prevent ambiguity in what they express [2, 6]. While machine understanding of unannotated text may become feasible [3], researchers are working to develop practical interfaces for human disambiguation of Web content [4]. To investigate methods of resolving one of the more difficult kinds of ambiguity, we conducted an experiment in which subjects disambiguated English sentences that contained syntactically ambiguous quantification [5].

METHOD

We selected 25 sentences from the Web (a small sample designed to encourage completion in an online, unmonitored testing environment). For each sentence, we identified two possible meanings and wrote a pair of paraphrases and an equivalent pair of truth conditions (situation descriptions) for them. For example, "Drinking almost always followed a dinner-party" had these restatements:

Paraphrases: (1) "Almost all drinking followed dinner-parties." (2) "Drinking followed almost all dinner-parties."

Truth conditions: (1) "In the activity diaries, 900 episodes of drinking were reported, and 875 of them followed dinner-parties." (2) "In the activity diaries, 900 dinner-parties were reported, and drinking followed 875 of them."

We asked some subjects (for method comparison) to choose between the paraphrases *or* between the truth conditions, and others (for consistency measurement) to choose *both* a paraphrase *and* a truth condition for each sentence. These two-task subjects might see the equivalent restatements in the same or in the opposite order.

We recruited 386 subjects: 208 through a Web contracting service [1], paid $0.75 each; and 178 through Internet discussion groups on language and writing, unpaid.

The ability to read and write English was the only participation requirement; 88% of the subjects had English as a native language. Subjects had opportunities to give us comments after each trial, after each block of 5 trials, and at the end of the experiment.

RESULTS

Satisfaction

Satisfaction was measured both by questionnaire responses, which indicated moderate satisfaction for all subjects (on three dimensions: ease, interest, and usefulness), and by completion rate. There were slight differences in satisfaction favoring paraphrasal over truth-conditional disambiguation and one-task over two-task conditions. For example, 90% of one-task subjects, compared with only 83% of two-task subjects, completed the experiment ($p < 0.04$).

Consistency, Speed, and Agreement

The choices made by a two-task subject in a trial were *consistent* if the chosen truth condition was equivalent to the chosen paraphrase. Choices were consistent in 82% of the trials, regardless of whether the paraphrasal or the truth-conditional task appeared first. But opposite-order trials (with the first paraphrase equivalent to the second truth

condition and vice versa) showed less consistency (76%) than same-order trials (86%). Of 159 subjects whose consistency rates differed between same- and opposite-order trials, 69% (109) were less consistent on opposite-order trials (two-tailed p < 0.00001).

The median time to perform a disambiguation was 20 seconds on one-task trials and 31 seconds on two-task trials. Truth-conditional selection typically took 23 percent longer than paraphrasal selection, perhaps because of the greater length and complexity of the truth conditions. Overall, the speed of disambiguation increased with experience.

The fastest subject to achieve 100% consistency finished in a total of 709 seconds. Others achieved 90% consistency in about 500 seconds, or 20 seconds per trial (see Figure 1).
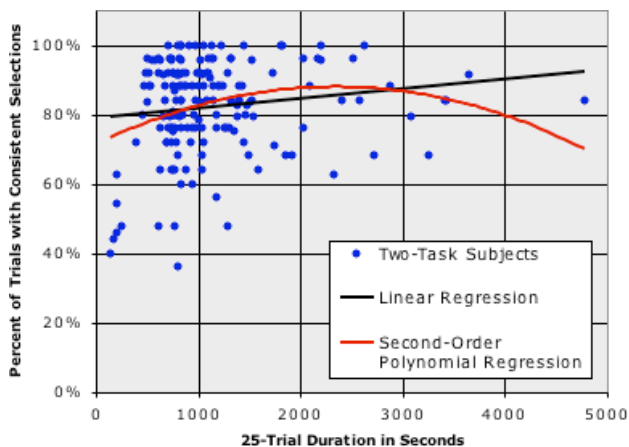

Figure 1. Consistency by Duration

Insofar as the majority correctly guesses intended meanings, the size of the majority is a measure of the subjects' collective success. We define a *method-majority* choice as the choice made by the majority of subjects (in all treatment groups) who disambiguated the same sentence with the same method in any trial. Of 13,859 choices made by all subjects, 77% were method-majority choices. This proportion was larger for paraphrasal selection (79%) than for truth-conditional selection (75%). Paraphrasing was the better method (it had higher method-majority rates) for 223 subjects, while truth-conditional selection was better for only 116 subjects ($p < 0.00000001$).

Subsample Analysis

By most measures, the unpaid volunteers performed better than the paid subjects. Of 79 two-task volunteers, 42 were more consistent than the overall median, vs. 37 of 95 paid subjects (2-tailed p = 0.0608). Of 178 volunteers, 87 made more than 1 comment, vs. 45 out of 208 paid subjects (2-tailed p < 0.0002). However, volunteers took longer: 84 of 178 volunteers took more than the overall median time to finish, vs. 52 of 208 paid subjects (2-tailed p < 0.0002).

Native and non-native speakers of English differed most strikingly in the disambiguation method that worked better for them. Most native speakers (202 of 340) agreed more often with the majority when using the paraphrasal method, but most (25 of 45) non-native speakers did so when using the truth-conditional method (2-tailed p = 0.0561). The truth conditions' emphasis on numerical rather than verbal reasoning may explain some of this difference.

DISCUSSION

One-task subjects resolved ambiguities in 15-25 seconds, with approximately 80% inter-method consistency and 80% majority agreement. Volunteers performed even better than paid subjects, reaching 99% agreement on the most consensual sentence. Many subjects, particularly in the volunteer subsample, described the disambiguation tasks as both challenging and enjoyable.

Our subjects guessed others' intended meanings, with no context but with the opportunity to choose between carefully crafted restatements. In future experiments, we intend to study disambiguation by authors, rather than readers, with more scalable methods of interactive disambiguation. We surmise that authors will be motivated to limit their ambiguity, just as our volunteers demonstrated their enthusiasm for disambiguation. Thus, we anticipate that the barriers to author disambiguation will be more technical than motivational. Our focus will be on developing methods that help motivated authors to recognize and reduce ambiguity.

REFERENCES

[1] Amazon.com, "Amazon Mechanical Turk" (Web site), 2007; http://www.mturk.com/mturk/welcome.

[2] Arnold, J. E., Wasow, T., Asudeh, A., and Alrenga, P., "Avoiding Attachment Ambiguities: The Role of Constituent Ordering", *Journal of Memory and Language*, 51, 2004, 55-70; http://www-csli.stanford.edu/~wasow/AWAA_final.pdf.

[3] Etzioni, O., Banko, M., and Cafarella, M. J., "Machine Reading", 2007 AAAI Spring Symposium on Machine Reading, 2007; http://turing.cs.washington.edu/papers/SS06EtzioniO.pdf.

[4] Kaufmann, E., and Bernstein, A., "How Useful are Natural Language Interfaces to the Semantic Web for Casual End-Users?", 6th International Symantic Web Conference (ISWC 2007), 2007; http://www.ifi.uzh.ch/ddis/staff/goehring/btw/files/Kaufmann_Bernstein_ISWC2007.pdf.

[5] Pool, J., and Colowick, S. M. (in press), "Disambiguating for the Web: A Test of Two Methods," *Proc. 4th Intl. Conf. on Knowledge Capture* (ACM Press, 2007); http://http://turing.cs.washington.edu/papers/disambweb.pdf.

[6] Wasow, T., Perfors, A., and Beaver, D., "The Puzzle of Ambiguity", in *Morphology and the Web of Grammar: Essays in Memory of Steven G. Lapointe*, ed. O. Orgun and P. Sells (Stanford: CSLI Publications, 2005); http://montague.stanford.edu/~dib/Publications/lapointe_paper_9-4.pdf.