Identifying words to explain to a reader: A preliminary study

Greg Aist

Language Technologies Institute, Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, Pennsylvania 15213
(412) 521-2862
aist@cs.cmu.edu
For further information see http://www.cs.cmu.edu/~aist/cv.html

Our goal: Help students learn vocabulary by explaining unfamiliar words

The central idea behind this paper is a familiar one to parents and teachers: While a child is reading a story, explain unfamiliar words. Project LISTEN's Reading Tutor (Mostow & Aist CALICO 1999) listens to children read aloud, using human-narrated stories to help kids learn to read. We would like to extend the capability of providing explanations for words to this automated reading tutor.

Initial solution: Fully automated annotations

In previous work we tried to automatically annotate text with semantic assistance on words. We augmented text with short "factoids" about words such as "cheetah can be a kind of cat. Is it here?" We used WordNet (Fellbaum 1998) to retrieve synonyms (assistance is like aid), hypernyms (cheetah is a kind of cat), and antonyms (beautiful is the opposite of ugly), and focused on the "low-hanging fruit" of words with only one or two senses.
During design, development and deployment of these automatic annotations, we uncovered a number of challenges to our automated approach.

- Appropriate language. Comprehensive language resources aimed at adults may not be appropriate for children because of "improper" language, too-difficult words, or archaic vocabulary.
- Multiple senses. We intended to skirt the problem of multiple senses by looking only at words with few senses, and presenting the remaining word sense disambiguation problem as an exercise to the reader. (In the hope that telling the difference between e.g. *elephant* the animal and *elephant* the political symbol would be easy enough to leave to the student.) Such phrases were sometimes confusing, and unfortunately the computer never gave the "answer" of what the correct sense for the word was in this context.
- Textual context. For some words it may be fine to explain them the same way regardless of the textual context in which they appear (e.g. *asparagus*, *aardvark*). However, some uses of a word are better explained by one synonym rather than others. Even harder, a word may be best explained by analogy with a word that is *not* a synonym, but a functionally related word. For example, one teacher in one of the schools we work with explained the word *slate* in a story about frontier schools in early America as being like a *chalkboard* -- you write on it in school. (In WordNet, *slate* and *chalkboard* are most closely related through *artifact*, fairly high up in the ontology.)

Together, these problems led us to seek instead a computer-assisted (but human-controlled) solution.

Revised solution: Elicit, capture, and utilize explanations from human expert

We would like to *elicit* explanations from a human expert, *capture* the explanations as narrated text, and *utilize* these explanations during assisted reading. Our plan for implementing these three steps is as follows.
To *elicit* explanations, we would like the computer to suggest words or phrases to annotate, but allow the expert to select words or phrases as well.
To *capture* explanations, we plan to rely on the computer tutor's authoring mode (Mostow & Aist USPTO 1999, Mostow & Aist AAAI 1999) which uses speech recognition to capture quality-controlled narrations of a typed-in story.
To *utilize* explanations during assisted reading, we will display the explanations as dynamic text -- sentences which are displayed on top of the original text, as commentary. The student will read the dynamic text aloud, with the computer's help. (We used this approach for the previous fully automated explanations.)

Explanations will be provided either on demand, or when the computer decides an explanation is warranted.

In the remainder of this paper we focus on one aspect of the first step: how to select words or phrases to suggest for annotation.

With unlimited time, money, and patience, we could employ a human expert to explain every word and phrase in every story used by the computer tutor. In practice some words are more important to explain -- because they are less likely to be known, or because they will be more useful in future. Also, the computer tutor will have to decide which explanations to present, so we want to capture expert judgement of *which* words should be explained.

In order to speed up the process of eliciting explanations, we would like the computer tutor's authoring tool to suggest words or phrases to annotate. How can we automatically suggest words to annotate? Eventually we would like to construct or learn heuristics that identify words to explain in previously unseen text. As a prerequisite, we need to determine whether word selection is entirely particular to each person, or is roughly the same for different people.

We conducted a brief exploration into what words or phrases people choose to explain. We had three people (male, native speakers of North American English, (at least) college graduates, but not experts in reading) annotate the poem "Paul Revere's Ride", by Henry Wadsworth Longfellow, a 19th century American poet. The poem is not included in its entirety due to its length (989 word tokens, or 840 not counting the articles *the* and *a*), but begins as follows:

Listen, my children, and you shall hear
Of the midnight ride of Paul Revere,
On the eighteenth of April, in Seventy-five;
Hardly a man is now alive
Who remembers that famous day and year.

He said to his friend, "If the British march
By land or sea from the town to-night,
Hang a lantern aloft in the belfry arch
Of the North Church tower as a signal light,--
One, if by land, and two, if by sea;
And I on the opposite shore will be,
Ready to ride and spread the alarm
Through every Middlesex village and farm
For the country folk to be up and to arm,"…

The instructions for the task were to annotate the story with one explanation for each word or phrase that the rater thinks should be explained to the reader. With the original poem quoted with '>' and annotations shown using an arrow (->):

>Listen, my children, and you shall hear
>Of the midnight ride of Paul Revere,
>On the eighteenth of April, in Seventy-five;
Seventy-five -> here, 1775
>Hardly a man is now alive
Hardly -> almost no
>Who remembers that famous day and year.
>
>He said to his friend, "If the British march
British -> from England
>By land or sea from the town to-night,
>Hang a lantern aloft in the belfry arch
aloft -> up
belfry -> church bell tower
arch -> [picture of arch]
>Of the North Church tower as a signal light,--
North Church -> an old church in Boston
signal -> like a sign
>One, if by land, and two, if by sea;
>And I on the opposite shore will be,
opposite -> across
>Ready to ride and spread the alarm
alarm -> warning
>Through every Middlesex village and farm
Middlesex -> the county that Boston is in
>For the country folk to be up and to arm,"
to arm -> to get their weapons

For analysis, we considered each word token a separate response, coded **1** if contained in an annotation, and **0** otherwise. We summarize the coding below:

Rater	Words coded **1** out of all words	Words coded **1** out of words w/out *the, a*
J	16% (160/989)	19% (159/840)
A	3% (26/989)	3% (26/840)
G	6% (59/989)	7% (59/840)

Pairwise interrater reliabilities using kappa (Carletta 1996):

	J	A	G
J	-	0.178 all words 0.172 w/out *the, a*	0.405 all words 0.397 w/out *the, a*
A	-	-	0.426 all words 0.422 w/out *the, a*
G	-	-	-

All values of kappa were significantly greater than zero, indicating agreement greater than chance. For reference, values of kappa below 0.40 show poor agreement, values between 0.40 and 0.75 show fair agreement, and values above 0.75 show excellent agreement (SPSS 1999).

Why is there reasonable agreement among two of the three pairs (J-G and A-G) but not between J and A? J and A differed greatly in the rate at which they coded words as needing explanation. Also, A commented that the instructions were unclear as to who was the intended reader of the annotated story. Finally, none of the three raters were experts in reading, although they did possess varying degrees of expertise in computer science and linguistics.

For these data, Cronbach's alpha was 0.6102 (0.6684 standardized item alpha) including all words and 0.6073 (0.6630 standardized item alpha) without the articles *the* and *a*. Desirable values of alpha vary, but a value of 0.7 or above commonly indicates a highly reliable test (Yu n.d.). We conclude that the coding task of annotating text with explanations of words or phrases shows promise as a replicable task, but needs to be revised and clarified. After iteration, we intend to test this coding scheme with certified elementary teachers using non-fiction texts.

References

Carletta, Jean. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2), 249-254. http://www.iccs.informatics.ed.ac.uk/~jeanc/squib.ps

Fellbaum, Christiane. 1998. WordNet: An electronic lexical database. Cambridge MA: MIT Press. Searchable index for WordNet 1.6 available at http://www.cogsci.princeton.edu/cgi-bin/webwn

USPTO99 Mostow, J. and Aist, G. Reading and Pronunciation Tutor. United States Patent No. 5,920,838. Filed June 2, 1997; issued July 6, 1999. US Patent and Trademark Office.

AAAI99 Mostow, J. and Aist, G. Authoring New Material in a Reading Tutor that Listens. Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99), Orlando, FL, July 1999, pp. 918-919. In the refereed Intelligent Systems Demonstration track. Also presented at 37th Annual Meeting of the Association for Computational Linguistics (ACL'99), College Park, MD, June, 1999.

CALICO99 Mostow, J. and Aist, G. Giving Help and Praise in a Reading Tutor with Imperfect Listening: Because Automated Speech Recognition Means Never Being Able to Say You're Certain. CALICO Journal16:3, 407-424. Special issue (M. Holland, Ed.), Tutors that Listen: Speech recognition for Language Learning, 1999.

SPSS. 1999. SPSS® Base 9.0 Applications Guide. Chicago IL: SPSS. See also company web site at http://www.spss.com

Yu, Alex. n.d. Using SAS for Item Analysis and Test Construction. Downloaded from the World Wide Web on January 17, 2000.
http://seamonkey.ed.asu.edu/~alex/teaching/assessment/alpha.html