

Eyetracking data can improve language technology and help readers

October 24 2016



Credit: University of Copenhagen

New research from the University of Copenhagen shows that recordings of gaze data - within a few seconds - can reveal whether a word causes a reader problems. This insight could be used to alleviate reading problems with software that offer translations of difficult words or suggest easier texts as soon as readers experience problems. The technology may thus have a significant impact on the educational system, particularly because gaze data can now be recorded with ordinary mobile phones and tablets.

When we read a [text](#), our eyes fixate repeatedly on words or sentences that we find difficult. This becomes clear when a computer is fitted with eyetracking technology which records eye movements. Linguists have

for a long time used the technology to study what happens in our minds when we read in order to develop general models of e.g. text difficulty. However, these types of studies require a lot of data and resources because researchers must test representative groups of words on as many readers as possible in order to say how certain types of words affect readers in general.

PhD Sigrid Klerke has adopted a different approach: rather than attempting to construct a general model that comprises many people and many types of texts, she has developed a model that is tailored to the individual reader; the model is based on the reader's [eye movements](#) the moment he or she reads a random text:

"Based on a few seconds of recorded gaze data from the eyetracking system, my model can assess whether the word you are looking at right now is a word that you find difficult. This information could, for example, be used to automatically suggest an explanatory note or a translation. The model has thus learned to note when you fixate on text in a characteristic pattern which we could not have described in advance," explains PhD Sigrid Klerke who has just defended her PhD thesis 'Glimpsed - improving [natural language processing](#) with gaze data' on how gaze data can be used to improve technology such as machine translation and automatic text simplification.

Resource-efficient technology

Compared to other types of language technology, Sigrid Klerke's system has the advantage that it does not require textual annotation:

"Within my field, language technology, we spend surprisingly many resources on hiring language experts to annotate texts as an important part of developing these technologies - e.g. by adding information about which part of speech a word belongs to or about which words can be

omitted. But the reader's gaze may be seen as a kind of annotation in real time, containing information that we just are beginning to understand how to use. Also, it is much faster getting someone to simply read a text than hiring experts to annotate the same text. What we do with our eyes when reading is by no means chance," says Sigrid Klerke and adds:

"No matter which text the system encounters, it will be able to assess how difficult the reader finds the text. It does not even have to consider which words the text contains, but can rely solely on the feedback from the reader's gaze. This means that we do not have to pay experts to annotate the texts or depend on general models of textual difficulty. Instead we utilize all the data that are generated while people are reading to make reading easier. As far as I know, this has never been done before."

Education software and Google Translate

An eyetracking-based reading support system as the one Sigrid Klerke has made possible will be in high demand among a number of commercial players. And within the education system, there will be a number of obvious applications:

"Inasmuch as eyetracking now can be built into mobile phones and tablets, it will be fairly easy to also install software that can assist learners when they are reading -for instance by adjusting the text's difficulty or suggesting texts which contain a certain type of words that the learner finds difficult and therefore needs to practice," suggests Sigrid Klerke.

"When Google and other major companies can begin to access user gaze data via mobile phones and tablets, they can use the feedback to improve their systems; if, for instance, a sufficient number of Danes fixate in the same pattern on the same word in a text Google Translate has translated

from English into Danish, this information may automatically be fed into their translation systems as an indication that the translation might be faulty. Google has so many users that this will generate astonishing amounts of useful data."

Sigrid Klerke also believes that eyetracking software could be used to prove that a person has read a document that she is obligated to read, and this will be of legal interest.

Provided by University of Copenhagen

Citation: Eyetracking data can improve language technology and help readers (2016, October 24) retrieved 16 June 2024 from <https://phys.org/news/2016-10-eyetracking-language-technology-readers.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.