

Using the Hypothes.is web annotation tool for neologism collection

Erin McKean

Wordnik

erin@wordnik.com

Abstract

Dictionary citation collection programs (sometimes called 'reading programs') involving both dedicated amateurs and paid professionals are not new, but have often required either cumbersome marking of print materials or creation of paper slips or access to private computer systems specific to individual projects. However, given the development and adoption of open standards for web annotation, citation collection by readers in and outside of dictionary programs can now be done easily without expensive proprietary tools or resorting to paper slips.

In this paper, we'll give an overview of Wordnik's reading program (currently in beta), which uses the free and open-source Hypothes.is web annotation tool to select, tag, and share citations from the open web directly for use on Wordnik.com. Using the Hypothes.is API, it is possible to import user-generated citations and their accompanying metadata directly into editorial workflows, including importing into KWIC corpora or other databases.

Since Wordnik is a radically inclusive dictionary (all words are eligible for inclusion), we'll discuss how this approach influences readers' marking of terms, and whether terms selected by readers are more likely to be typical neologisms (newly-coined words) or words overlooked by traditional dictionaries (e.g. jargon, slang, nonce, or other low-frequency words).

Keywords

dictionary users, web annotation, neologisms, hypothes.is, free-range definition