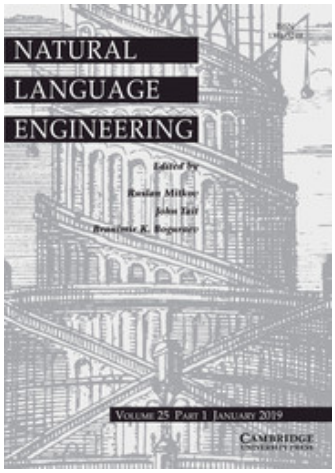

Identifying signs of syntactic complexity for rule-based sentence simplification

Natural Language Engineering, Volume 25, Issue 1

RICHARD EVANS, CONSTANTIN ORĂSAN

DOI: 10.1017/S1351324918000384

Published online: 31 October 2018, pp. 69-119

Print publication: January 2019

Abstract

This article presents a new method to automatically simplify English sentences. The approach is designed to reduce the number of compound clauses and nominally bound relative clauses in input sentences. The article provides an overview of a corpus annotated with information about various explicit signs of syntactic complexity and describes the two major components of a sentence simplification method that works by exploiting information on the signs occurring in the sentences of a text. The first component is a sign tagger which automatically classifies signs in accordance with the annotation scheme used to annotate the corpus. The second component is an iterative rule-based sentence transformation tool. Exploiting the sign tagger in conjunction with other NLP components, the sentence transformation tool automatically rewrites long sentences containing compound clauses and nominally bound relative clauses as sequences of shorter single-clause sentences. Evaluation of the different components reveals acceptable performance in rewriting sentences containing compound clauses but less accuracy when rewriting sentences containing nominally bound relative clauses. A detailed error analysis revealed that the major sources of error include inaccurate sign tagging, the relatively limited coverage of the rules used to rewrite sentences, and an inability to discriminate between various subtypes of clause coordination. Despite this, the system performed well in comparison with two baselines. This finding was reinforced by automatic estimations of the readability of system output and by surveys of readers' opinions about the accuracy, accessibility, and meaning of this output.