Abstract

Resolving prepositional-phrase (PP) attachment ambiguity is a challenging task in natural language processing. Unlike English language, researchers has paid little attention to address this problem in Arabic language. In this study, we use word collocation data derived from a large Arabic corpus to predict the most likely interpretation of potentially ambiguous PP-attachment phrases. We administered an empirical study in which human participants were presented with Arabic text involving potential PP-attachment ambiguity and their task was to judge whether the PP is attached to the preceding noun (low attachment) or verb (high attachment), or it is unclear. This exercise was used to collect a small-size labelled corpus of 50 examples (= 5 prepositions x 10 phrases). Subsequently, this labeled corpus was analysed to derive rules based on words collocational frequencies obtained from sketch engine operated on arTenTen12 corpus. Finally, the derived rules were validated using human judgment on unseen examples which were not used during the rules derivation step. We achieve 83% precision and 88% recall, which suggests that words collocation data generated by sketch engine can be used to resolve PP-attachment ambiguities.
References

18. R. Al-sabbagh and K. Elghamry, "A Web-based approach for Arabic PP-attachment.," in


Index Terms

Computer Science
Information Sciences

Keywords

Arabic word sketches, pp-attachment ambiguity, ambiguity resolution, arTenTen12 corpus, sketch engine