

Resumen

Resolución automática de la anáfora indirecta en el Español

Se describe la investigación para desarrollar un método de resolución de la anáfora indirecta en el Español. El método utiliza diccionarios de sinónimos y de escenarios, enriquecidos con el contexto lingüístico para poder determinar la presencia de la anáfora indirecta.

Al iniciar la investigación se estableció la necesidad de descubrir los marcadores que identificaban la presencia de la anáfora indirecta en los textos escritos; conforme avanzaba el análisis fue necesario profundizar en la relación existente entre la sintaxis, la semántica y la pragmática cuando participan activamente en el discurso textual; con el fin de integrar el contexto lingüístico como parte del discurso comunicativo se requirió la identificación del rol de las expresiones referenciales, marcadas por los determinantes en las frases nominales; finalmente el descubrimiento de la fuerte interrelación entre la referencia, correferencia y la anáfora permitió desarrollar un modelo de escenario que identifica más claramente la presencia de la anáfora indirecta.

Se presentan los fundamentos y resultados que prueban la factibilidad de tres hipótesis: primera “la anáfora indirecta posee los mismos marcadores que la referencia y la correferencia en el discurso, la diferencia estriba en el tipo de relación, el método de inferencia y la información necesaria para resolverlas”; segunda “la anáfora indirecta se presenta si, y sólo si, no existe correferencia”; y tercera “la referencia se presenta si, y sólo si, no existen correferencia y anáfora indirecta”. La prueba manual de las hipótesis permitió modelar e implantar un sistema que detecta y resuelve la anáfora indirecta con un prototipo y diccionarios de sinónimos y escenarios validados manualmente. Para trabajar con texto libre se limita a la anáfora indirecta nominal debido a que depende de la información del diccionario de escenarios y actualmente no existe un diccionario de relaciones verbal-nominal en el Español.

Abstract

Automatic resolution of indirect anaphora in Spanish

This thesis describes the research to develop a method for indirect anaphora resolution in Spanish. This method uses a scenario dictionary enriched with the linguistic context to determine indirect anaphora presence.

At the beginning of the research the necessity to discover the markers, that identified indirect anaphora presence, was outlined; according to the analysis advance it was necessary to deep in syntax, semantic and pragmatic relation when actively participate in textual discourse; in order to integrate the linguistic context as part of the communicative discourse identification of reference expressions, marked by determinants in nominal phrases, was required; finally, the discovery of the strong interrelation between reference, coreference and indirect anaphora let developed an enlarged scenario model that more clearly identified indirect anaphora presence.

Fundaments and results that evidence feasibility of three hypotheses are presented: first “indirect anaphora has same markers than reference and coreference in discourse, the difference lie on the type of relation, inference method and information necessary to resolve them”; second “indirect anaphora is present if, and only if, there is not coreference”; and third “reference is present if, and only if, there is not indirect anaphora”. Manual test of these hypotheses let model and implement a system that detects and resolve indirect anaphora with a prototype and dictionaries handly verified. The work with free text indirect anaphora resolution is limited to nominal expressions because it depends on scenario dictionary, and actually there is not a verbal to nominal relations dictionary in Spanish.

