

Klassifizierung und Etikettierung der Nomina in Korpora des gesprochenen Italienisch

Die vorliegende Arbeit verfolgte das Ziel, ein geeignetes Tagset zur morphosyntaktischen Annotation der Nomina in Korpora des gesprochenen Italienisch zu erstellen. Dazu wurde zunächst die Einteilung der Nomina in drei ausgewählten italienischen Wörterbüchern und in drei ausgewählten italienischen Grammatiken untersucht. Anschließend wurde ein Überblick darüber gegeben, wie Substantive nach den EAGLES-Richtlinien, das heißt nach den internationalen Standards zu annotieren sind. Der nächste Schritt bestand in der Analyse der Klassifizierung und Etikettierung der Nomina in allen bislang morphosyntaktisch annotierten Korpora des gesprochenen Italienisch, zu denen das *Corpus del lessico di frequenza dell'italiano parlato* oder LIP-Korpus, das *Archivio del Parlato Italiano* oder API-Korpus sowie die italienische Komponente des C-ORAL-ROM-Korpus zählen. Außerdem wurde die Einteilung der Nomina auch in zwei englischen Korpora gesprochener Sprache, im *British National Corpus* und in der britischen Komponente des *International Corpus of English*, analysiert.

Das in Kapitel 4 vorgeschlagene Tagset zur Etikettierung der Nomina in Korpora des gesprochenen Italienisch ist von einem logischen Aufbau gekennzeichnet, das heißt, die Beziehungen zwischen den einzelnen Unterklassen können mittels eines Hierarchiebaumes dargestellt werden. In diesem Tagset werden Substantive zunächst in die beiden semantischen Unterklassen ‚Eigennamen‘ und ‚Gattungsnamen‘ unterteilt. Außerdem wird die Klasse der Gattungsnamen in insgesamt neun morphologische Subklassen eingeteilt; die Klasse der Eigennamen wird hingegen nicht subklassifiziert. Die einzelnen Komponenten von Mehrwortlexemen sind nach dem eigenen Vorschlag als eine lexikalische Einheit zu betrachten, wobei der gesamten Sequenz lediglich eine Etiketete zuzuordnen ist. Kontraktionen müssen in ihre einzelnen Bestandteile aufgespaltet und getrennt voneinander annotiert werden. Fremdwörter sind nach dem eigenen Vorschlag mit denselben Etiketten zu kennzeichnen wie Nomina, die keine Fremdwörter darstellen, um auf diese Weise Probleme bei der Abgrenzung fremdsprachiger Nomina von italienischen Nomina zu vermeiden.