

Automatische Vernetzung von Verben mit kontrastiven Valenzstrukturen in UBY

Judith Eckle-Kohler und Christian M. Meyer

Technische Universität Darmstadt
Ubiquitous Knowledge Processing (UKP) Lab
Hochschulstraße 10
64289 Darmstadt
<http://www.ukp.tu-darmstadt.de>

In der vorliegenden Arbeit stellen wir ein automatisches Framework vor, mit dem zweisprachige Listen von Verben mit kontrastiven Valenzstrukturen in den verwandten Sprachen Deutsch und Englisch aus der lexikalischen Ressource UBY (Gurevych et al., 2012) extrahiert werden können. Kontrastive Valenzstrukturen sind dadurch gekennzeichnet, dass sie sehr ähnliche Satzstrukturen beschreiben und sich nur in einem Merkmal – dem Kontrast – unterscheiden. Beispielsweise haben Verben mit einer nicht-finiten Verbergänzung weitgehend übereinstimmende Valenzstrukturen im Deutschen und Englischen. Ein Kontrast kann jedoch bei der Form der nicht-finiten Verbergänzung vorliegen: *zu*-Infinitiv im Deutschen versus *to*-Infinitiv oder Gerundium im Englischen.

Solche Verblisten für ein gegebenes Sprachpaar manuell zu ermitteln, ist sehr aufwendig. Fischer (1997) betrachtet in seiner kontrastiven Analyse zur Verbvalenz in Deutsch und Englisch beispielsweise nur 17 Verben. Wir schlagen daher ein automatisches Framework vor, um möglichst umfassende Listen von Verben mit kontrastiven Argumentstrukturen zu extrahieren. Wir nutzen dazu die lexikalische Ressource UBY, in der zahlreiche Wörterbücher und Wortnetze in einem einheitlichen Datenmodell repräsentiert werden. Ein wesentlicher Teil dieses Datenmodells modelliert Subkategorisierungsrahmen aus IMSlex-Subcat und GermaNet für Deutsch sowie aus VerbNet für Englisch in einer einheitlichen Weise (Eckle-Kohler, 2012).

In der vorliegenden Arbeit entwickeln wir ein automatisches, auf der Ähnlichkeit von Subkategorisierungsrahmen basierendes Verfahren zur Verknüpfung äquivalenter Verbbedeutungen über die Grenzen dieser drei Lexika hinweg. Für das eingangs erwähnte Beispiel betrachten wir Verben, die im Deutschen mit *zu*-Infinitiv verwendet werden können und extrahieren automatisch englische Entsprechungen, die mit Gerundium verwendet werden und daher Beispiele für diesen spezifischen Kontrast sind. Das vorgestellte Framework erlaubt ähnliche Analysen für andere Kontraste, etwa zur Nutzung von Korrelaten bei deutschen Komplementsätzen, für die es im Englischen keine Entsprechung gibt. In einer manuellen Auswertung der verknüpften Verbbedeutungen durch zwei unabhängige Annotatoren ergab sich ein korrektes Ergebnis in 90% (IMSlex-Subcat – GermaNet) bzw. 88% (IMSlex-Subcat – VerbNet) der Fälle.

Derartige Verblisten für kontrastive Valenzstrukturen stellen eine interessante Ressource für die kontrastive Linguistik und die Zweitspracherwerbsforschung dar. Eine mögliche praktische Anwendung sehen wir in der Erzeugung von kontrastivem Übungsmaterial zur korrekten Verwendung von englischen und deutschen Verben.

Bibliographie

Judith Eckle-Kohler und Iryna Gurevych (2012): Subcat-LMF: Fleshing out a standardized format for subcategorization frame interoperability. In: *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 550–560, Avignon, France.

Klaus Fischer (1997): *German-Englisch verb valency: a contrastive analysis*. Tübingen: Narr (= Tübinger Beiträge zur Linguistik 422).

Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, und Christian Wirth (2012): UBY – A Large-Scale Unified Lexical-Semantic Resource Based on LMF. In: *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 580–590. Avignon, Frankreich.