

Kolloquium, 29. Mai 2017

Dayyan Smith

Disambiguierung eines japanischen Aspektmarkers mithilfe von Parallelkorpora

Vortragender: Korbinian Schmidhuber

Betreuerin: Annemarie Friedrich

Das Ziel von Korbinians Bachelorarbeit ist es einen Klassifikator zu erstellen für die Disambiguierung eines japanischen Aspektmarkers mithilfe von Parallelkorpora (japanisch, englisch). Motiviert davon, dass regelbasierte Systeme in der Computerlinguistik oft nicht umsetzbar sind, da sie zu abstrakt sind, möchte Korbinian in seiner Bachelorarbeit auf ein beispelsbasiertes System setzen. Die *te iru* Verbform im Japanischen entspricht weitestgehend dem Englischen Progressive. Im Englischen wird die Verlaufsform durch das Progressive gebildet, während ein Zustand nicht durch das Progressive ausgedrückt werden kann.

Drei Korpora kommen zur Anwendung: Ein Wikipedia Korpus, bei dem der japanische Teil von Hand aus dem Englischen übersetzt wurde, ein “Basic Sentences” Korpus und ein Korpus bestehend aus Ausgaben von “Wachstum” auf englisch und japanisch. Für jeden Satz im Englischen gibt es in den Korpora einen Satz im Japanischen.

Im ersten Schritt wurden Teilkorpora erstellt, die “te iru” Konstruktionen beinhalten. Danach mussten für zwei der Korpora noch die Verben aligniert werden, wofür GIZA++ und fast_align zum Einsatz kamen. Der nächste Schritt ist die Anwendung und der Vergleich verschiedener Klassifikatoren für das Disambiguieren.

Bei der Alignierung sind Sprachpaare mit unterschiedlicher Wortreihenfolge grundlegend problematisch und führen zu schlechten Ergebnissen. Als weiteres Problem ist aufgefallen, dass es Konstellationen mit *te iru* gibt, die nicht einen Verlauf beschreiben, sondern das Ergebnis einer Handlung.

Wegen Abbruch der Bachelorarbeit liegen keine weiteren Ergebnisse vor.