

Übungsblatt 8

8.1 Distributionelle Ähnlichkeit

In Aufgabe 7.2 sollten für 15 Begriffspaare im Deutschen die englischen Entsprechungen gesucht und die WordNet-Verbindung der englischen Äquivalente untersucht werden. Die deutschen Ausgangsbegriffe waren:

Fahrrad: *Fahrzeug, Sport, Rad, Pedal, Moped*

Erdbeere: *Kuchen, Frucht, Sahne, Garten, süß*

Buch: *Seite, Roman, Autor, Bibliothek, lesen*

- a. Berechnen Sie für die Pfade, die Sie gefunden haben, die WordNet-Ähnlichkeit zwischen den englischen Ausdrücken nach der in den Vorlesungsfolien angegebenen Gleichung (kein Pfad zwischen w_1 und w_2 vorhanden: $\text{sim_wn}(w_1, w_2) = 0$).
- b. Es gibt ein Web-Interface, das semantische Ähnlichkeitswerte für englische Ausdrücke auf der Grundlage der Verteilung im BNC (British National Corpus) angibt (www.iccs.informatics.ed.ac.uk/~scottm/semantic_space_model.html). Ermitteln Sie (i) die Ähnlichkeitsrelationen für die jeweiligen Paare und lassen Sie sich (ii) die zehn semantisch ähnlichsten Wörter für jeden der 3 (englischen) Ausgangsbegriffe geben. Verwenden Sie in beiden Fällen die erste Einstellung (lemmatisiert, auf 100 MW Korpus).
- c. Vergleichen Sie die Ergebnisse für WN-Ähnlichkeit und distributionelle Ähnlichkeit. Wie gut und wie vollständig approximieren die beiden Maße einen intuitiven Begriff von semantischer Ähnlichkeit?

8.2 Verarbeitung gesprochener Sprache (Ergänzung zur nächsten Vorlesung)

- a. Lesen Sie aus dem CL-Handbuch (K.-U. Carstensen et al., Computerlinguistik und Sprachtechnologie, 3. Auflage 2010) den Artikel 3.3.2 (Sprachsynthese). Kopie in der Bibliothek.
- b. Der Artikel zeigt an einem Beispiel, dass für eine korrekte Aussprache die Kenntnis über die interne morphologische Struktur des Wortes relevant ist. Geben Sie (mindestens) zwei weitere (möglichst vom Typ her unterschiedliche) Beispiele, und kommentieren Sie sie kurz.
- c. Der Artikel führt (im Abschnitt über „concept-to-speech“) aus, dass gute Sprachsynthese eine gewisse Kenntnis über grammatische Struktur und Bedeutungsinformation voraussetzt, gibt aber keine konkreten Beispiele. Versuchen

Sie, zwei unterschiedliche Beispiele für die Abhängigkeit der Aussprache von Syntax und/oder Semantik zu finden, und kommentieren Sie sie!

- d. Die Firma Nuance, zurzeit fast Monopolist für Sprachsoftware, bietet ein Online-Interface für TTS an, unter <http://www.nuance.com/vocalizer5/flash/>. Wörter/Sätze werden eingetippt und mit einer vorher ausgewählten Stimme akustisch realisiert. Probieren Sie das System aus, indem Sie (zunächst eine deutsche Stimme wählen und dann) eine Reihe von Beispielen eintippen (Beispiele aus dem Artikel oder frei gewählte). Beschreiben Sie stichwortartig, welche Aspekte beim System gut/ weniger gut behandelt werden bzw. zu Fehlern führen.
- e. Geben Sie die Ausdrücke ein, die sie zu den Fragen b. und c. gefunden haben, und beschreiben Sie den Effekt. Wenn das System in allen Fällen korrekt verbalisiert, suchen Sie nach Alternativen, bei denen das System Fehler macht.