IDS | LEIBNIZ-INSTITUT FÜR DEUTSCHE SPRACHE

Syntaktische Analysen im Nähe-Distanz-Kontinuum

Sarah Broll

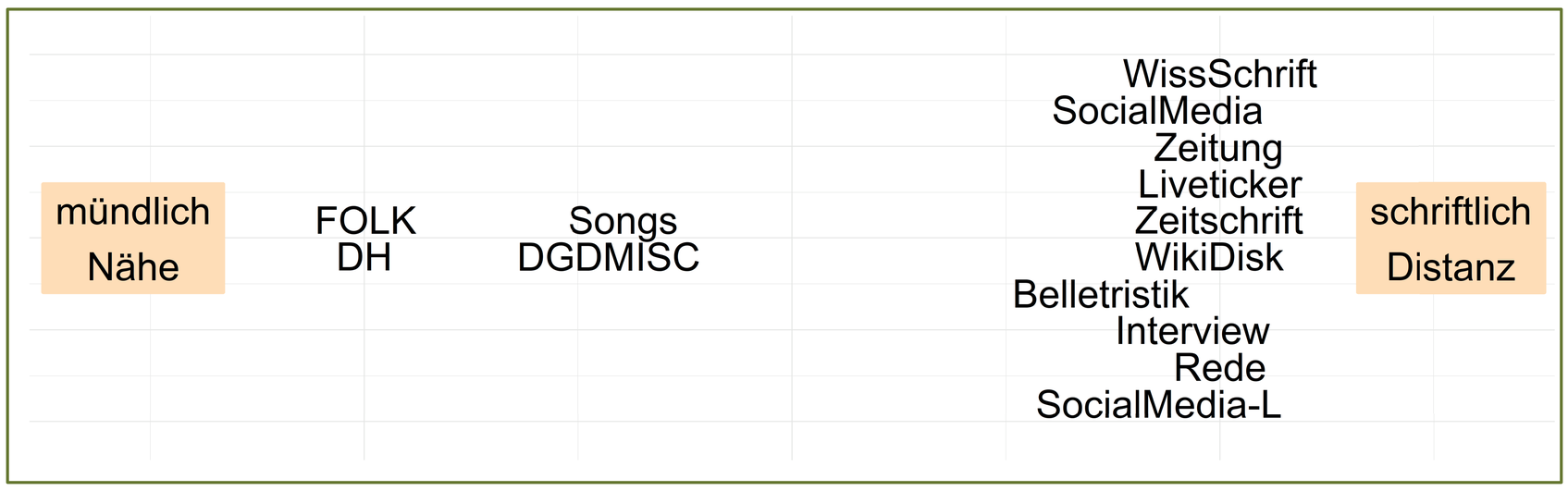
Internationales **Promovierenden-Netzwerk** des IDS

Anhand des Korpus ***COrLiCo*** (*Corpus for the Oral-Literate Continuum*), das Texte versammelt, die über das Nähe-Distanz-Kontinuum (Ágel und Hennig, 2006) hinweg verortet werden, soll die Ausprägung eines syntaktischen Phänomens in unterschiedlichen Textsorten untersucht werden.

Datengrundlage

Das Korpus ist ein ausgewogenes Korpus mit 120 Millionen Wort-Tokens, die zu ungefähr gleichen auf Teilen auf ca. 20 Textsorten aufgeteilt werden.

Eine Auswahl an Textsorten wurde anhand eines Random-Forest-Klassifikators mit einer Genauigkeit von 0,99429 (F1 = 0,9943) nach Konzeption klassifiziert.



Basierend auf den Daten aus Broll und Schneider (i. V.)

Dependenzannotation

Verschiedene **Optionen** für eine automatische Annotation von Dependenzrelationen:

- UDPipe
- Spacy

Keine Modelle für mündliche Texte!

- Eigenes Modell

Goldstandard annotieren!

Satzgrenzenerkennung

Problem: Einige Texte haben keine Satzgrenzenmarkierung!

Potentielle **Lösungen**:

- (Teilweise) Interpunktion einfügen

  Eigene interpretatorische Leistung!

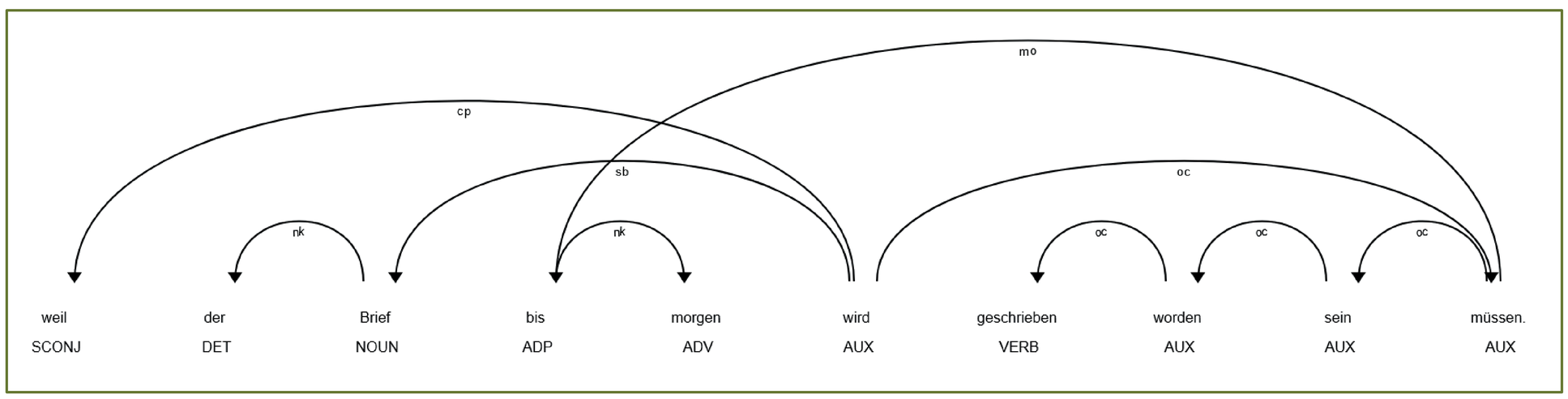
- Communication Units

Phänomenauswahl

Sprachplanungseinflüsse auf die Form eines Phänomens oder einer Phänomengruppe anhand von, z. B.:

- Komplexität des Verbalkomplex

Die Phänomenauswahl ist dabei abhängig von den Frequenzen in der Datengrundlage.



Satz von: https://grammis.ids-mannheim.de/systematische-grammatik/1522

Kontakt:
Sarah Broll
Abteilung Grammatik
Leibniz-Institut für Deutsche Sprache
Postfach 10 16 21
68016 Mannheim

Tel.: +49 621 1581-256
Fax: +49 621 1581-200
broll@ids-mannheim.de

Hausanschrift:
Leibniz-Institut für Deutsche Sprache
R 5, 6-13
68161 Mannheim

Tel.: +49 621 1581-0
Fax: +49 621 1581-200
info@ids-mannheim.de
www.ids-mannheim.de

© 2023 IDS Mannheim/ÖA

Leibniz Gemeinschaft

Literatur

Ágel, V. & Hennig, M. (2006). Theorie des Nähe- und Distanzsprechens. In V. Ágel & M. Hennig (Hrsg.), Grammatik aus Nähe und Distanz: Theorie und Praxis am Beispiel von Nähetexten 1650-2000 (S. 3-31). Tübingen: Niemeyer.

Broll, S. & Schneider, R. (i. V.). Empirische Verortung konzeptioneller Nähe/Mündlichkeit inner- und außerhalb schriftsprachlicher Korpora.