JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

MODEL UND EVIDENZ

UNI FREIBURG

„Ich gehe kurz Zigaretten holen" -

DISKURS berechnen mit WORD EMBEDDING

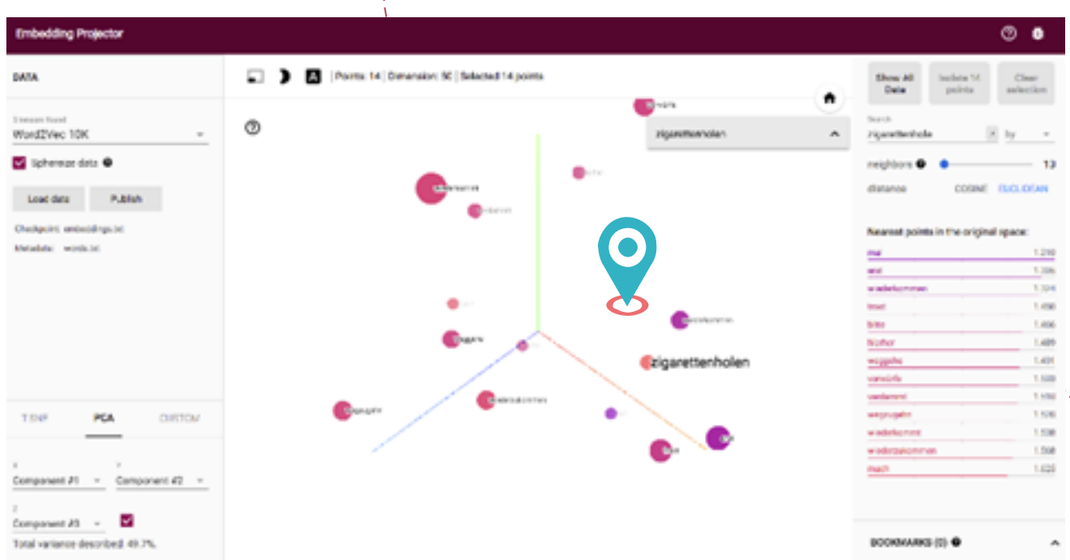
Dr. Zakharia Pourtskhvanidze
Institut für Empirische Sprachwissenschaft. Goethe-Universität Frankfurt

GRUNDANNAHMEN

(1) Die Diskurse lassen sich mithilfe von korpuslinguistischen Tools grundsätzlich berechnen. (Bubenhofer 2008);
(2) Die Satzhypostasen sind Sprachgebrauchsmuster und indizieren die spezifischen korrespondierenden Diskurse.

PROBLEMSTELLUNG

(1)
Überprüfung der Schlagwortfähigkeit der Satzhypostase
„Ich-gehe-nur-mal-kurz-Zigaretten-holen";
(2)
Ermittlung von korrespondierenden Diskurse im korpuslinguistischen verfahren (Word Embedding).



('warten', 0.8982756733894348), ('geblufft', 0.8894602656364441), ('weggehn', 0.888272762298584), ('satt', 0.8826419711112976), ('gerne', 0.8819628953933716), ('wiederzukommen', 0.8776798248291016), ('wiederkommen', 0.8767836093902588), ('endlich', 0.8742311000823975), ('wohin', 0.8716533780097961), ('woanders', 0.8713783025741577)

ERGEBNIS

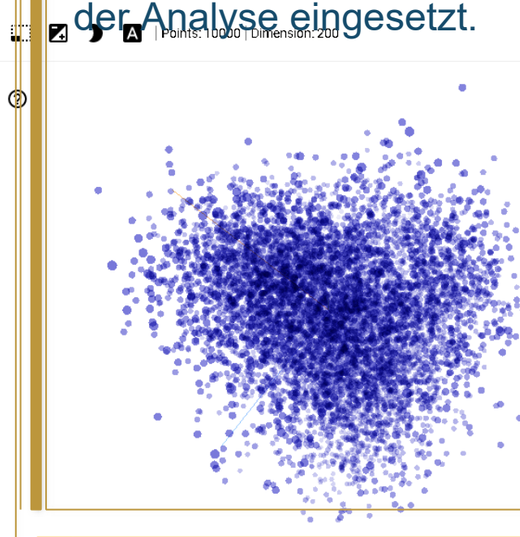
Die Konstruktion „Ich-gehe-nur-kurz-Zigaretten-holen" hat sich mit der konzeptuellen Bedeutung „weggehen ohne sich zu verabschieden und/oder Gründe zu nennen" im Beziehungs- bzw. Liebes-Diskurs verselbstständigt.

Die Konzepte wie {„weggehn", „bluff", „wieder(zu)kommen"} sind im Word-Embedding-Verfahren errechnet und stehen als korrespondierenden Diskurs-Schlagwörter.

ONTHOLOGIE DES ZIGARETTENHOLEN-DISKURSES IM SPRACHGEBRAUCH

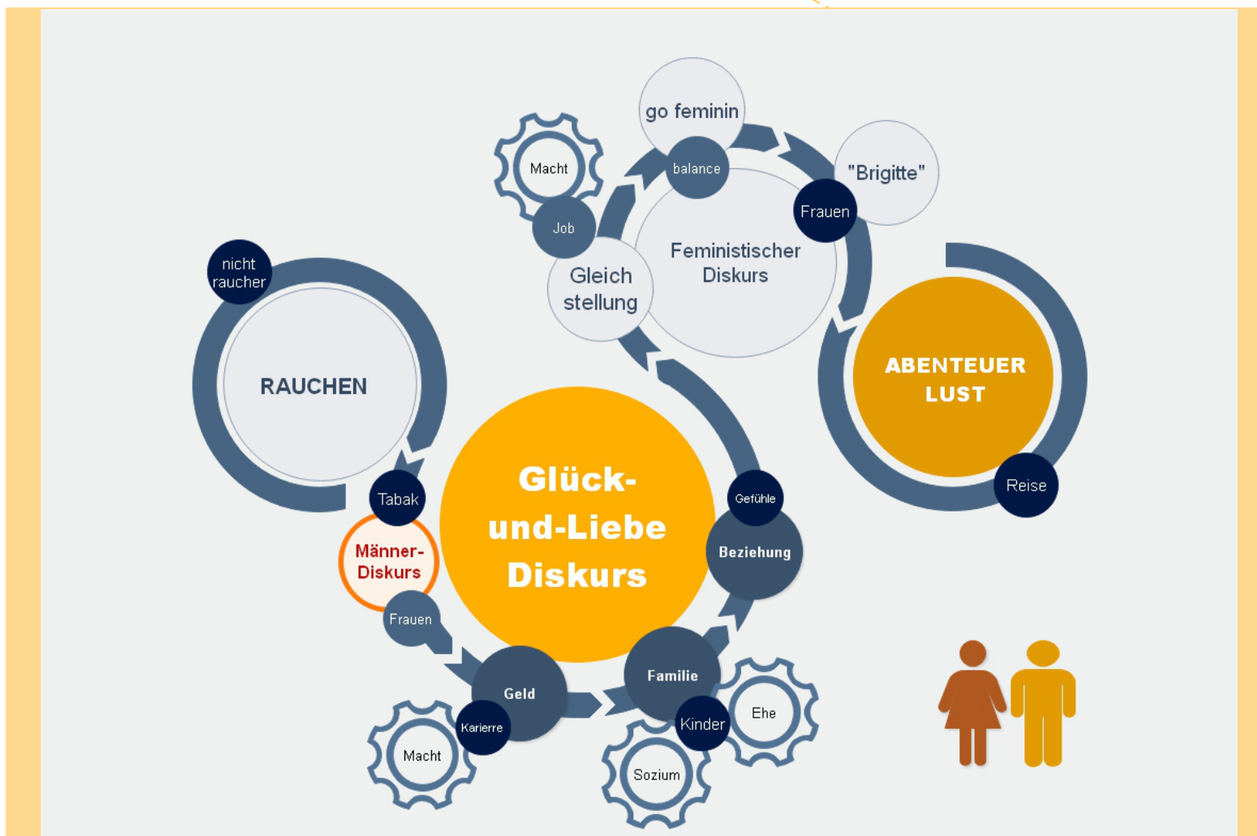
"ZIGARETTENHOLEN" zwischen WORD- und PHRASE EMBEDDING

Es wurde ein Webkorpus aus ca. 1Mio. Token verwendet. Im 50-dimensionalen Tensor-Raum wurden die Vektorendaten analysiert und 14 miteinander zusammenhängende Knoten isoliert. Die verschiedenen phrastischen Versionen des Satzes wurden in einem **Uni-Gram** „Zigarettenholen" umgewandelt und der Vektor dieses kumulativen Uni-Grams wurde in der Analyse eingesetzt.



Durch das Netzwerk-Modell, dass aufgrund der Word Vectors Analyse entsteht, lässt sich eine allgemeine Diskurs-Klassifizierung vollziehen, in dem die folgenden Diskurs schlag-wörter errechnet wurden:

1. '*warten*'
2. '*geblufft*'
3. '*weggehn*'



Bubenhofer, Noah. 2008. Diskurse berechnen? Wege zu einer korpuslinguistischen Diskursanalyse. In: Spitzmüller/Warnke. Methoden der Diskurslinguistik: sprachwissenschaftliche Zugänge zur transtextuellen Ebene. Berlin / New York: de Gruyter, 407-434.

pourtskhvanidze@em.uni-frankfurt.de