technische universität dortmund
Thomas Bartz thomas.bartz@tu-dortmund.de

UNIVERSITÄT MANNHEIM
Angelika Storrer astorrer@mail.uni-mannheim.de

Visualisierung von Frequenzverteilungen

Anwendungsbereich korpusbasierte Lexikographie

auf Basis von Wortformen

Kernkorpus 20

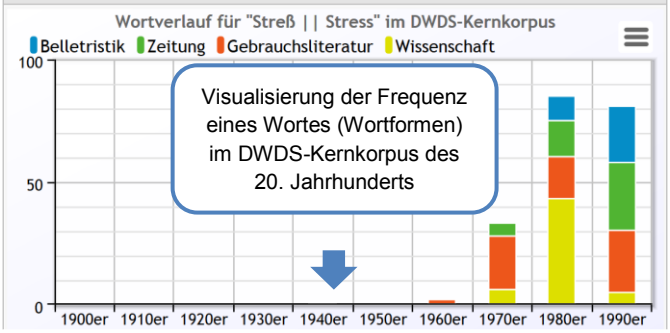
Treffer: **203**, davon anzeigbar: 115

Streß || Stress

KWiC | Datum ↓

1 1999 Abend war ich sehr müde - viel **Streß** auf Arbeit gehabt und Streit n
2 1999 ts ärgert es mich , daß ich unter **Streß** packen muß und dann wiede
3 1999 der menschlichen Affekte unter **Streß** .
4 1999 e Szene , in der Eliza unter dem **Streß** des Übens fast in Tränen aus

Wortverlauf (Basis DWDS-Kernkorpus)



Visualisierung der Frequenz eines Wortes (Wortformen) im DWDS-Kernkorpus des 20. Jahrhunderts

Etymologisches W

Nutzung der Erkenntnisse zur Frequenzverteilung für lexikographische Angaben

Streß

Streßm. '(gesundh des menschlichen Organismus durch große physische und psychische Belastungen', Übernahme (60er Jahre 20. Jh.) von gleichbed. amerik.-engl.*stress*, das zuerst 1936 von dem kanadischen Physiologen H. Selye als medizinischer Terminus im oben angegebenen Sinne verwendet wird.

auf Basis lexikalischer Einheiten

Kernkorpus 20

Treffer: **37**, davon anzeigbar: 36

zeitnah

KWiC | Datum ↓

1 1999 nde im Zusammenhang mit der **zeitnahen** Privatisierung der deutsc
2 1999 rechnen , " auf einen anderen **zeitnah** späteren Flug bzw. Zug " u
3 1986 er die filmische Gestaltung einer **zei** , in de
4 1986 ie **zeitnahe** Komödie um junge Mensc
5 1986 Dieser **zeitna** sc
6 1986 en deutschen Dramatiker für ein **zeitnahes** Bühnenwerk gestiftet wo
7 1986 Diese **zeitnahe** Liebesgeschichte um ein l
8 1986 das Stück doch auf seine Weise **zeitnah** : Die Revolution läßt die He

zeitnahe Privatisierung
zeitnahe Lebensgeschichte
zeitnah späterer Flug

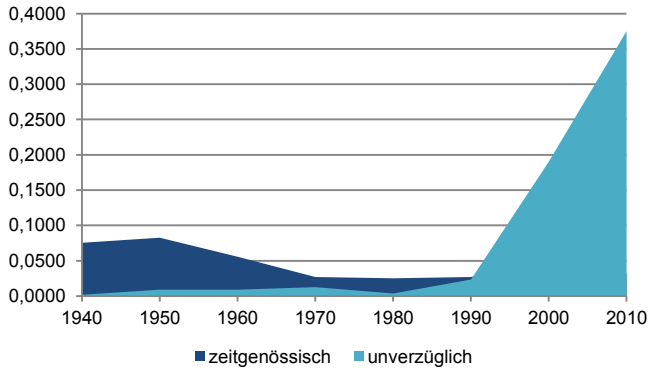
zeitnah
1. gegenwartsnah, zeitgemäß, zeitkritisch
2. schnell, unverzüglich, prompt

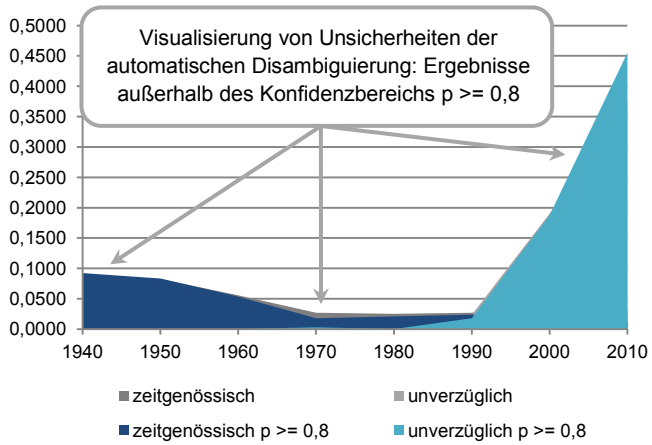
Belege

DWDS-Kernkorpus des 20. Jh.	ZEIT-Korpus
37	251

Automatische Disambiguierung
Verfahren: Latent-Dirichlet-Allocation (LDA, vgl. Blei et al. 2003)
Disambiguierung auf Basis des Wortkontextes, Kontext: 3 Sätze
Güte des Verfahrens: F_1: 0,88; NMI: 0,418

Visualisierung der Frequenz der Bedeutungen im ZEIT-Korpus





Visualisierung von Unsicherheiten der automatischen Disambiguierung: Ergebnisse außerhalb des Konfidenzbereichs p >= 0,8

Projekthintergrund

Korpus-basierte linguistische Recherche und Analyse mithilfe von Data-Mining

ZIEL
Verbesserung und Beschleunigung der korpus-basierten Recherche und Analyse

ERGEBNIS
Werkzeugkasten für die korpus-basierte computergestützte linguistische Recherche und Analyse

AUFGABEN
- Belegfilterung
- Disambiguierung
- Annotation

Welche Aufgaben mit welchen Methoden?

METHODEN
- Klassifizieren
- Clustering
- Aktives Lernen
- Hybride Ansätze

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
technische universität dortmund
UNIVERSITÄT MANNHEIM
EBERHARD KARLS UNIVERSITÄT TÜBINGEN
INSTITUT FÜR DEUTSCHE SPRACHE
berlin-brandenburgische AKADEMIE DER WISSENSCHAFTEN

http://www.kobra.tu-dortmund.de