

Auch von mir noch mal willkommen und schön, dass Sie da sind

# Diskursmarker

in  
**schriftlichem & akustischem**  
**Diskurs**

BACHELORVERTEIDIGUNG

Johanna Sacher, 4.2.2021



- In unserem Medienkonsum heute wird immer mehr gehört
- Streamingdienste: Hörbücher, Podcast, Plattformen wie YouTube
- und Sprachassistenten lesen auf Wunsch die Zeitung vor
- Für Konsumenten ist das praktisch
- Doch für Produzenten und Forscher stellt diese Veränderung ein Problem dar:
- Die meisten unserer Tools, z.B. Analyseprogramme und Schreibassistentensoftware, wurden für schriftliche Texte entwickelt, nicht für akustische
- schon intuitiv ist es so, dass wir anders reden, als wir schreiben, wir hören anders, als wir lesen
- Dieser Annahme gehe ich in meiner Arbeit auf den Grund
- Konkret habe ich am Beispiel von Diskursmarkern untersucht, wie sich akustischer und schriftlicher Diskurs unterscheiden
- Und meine Ergebnisse möchte ich Ihnen jetzt gerne vorstellen

Diese Arbeit liefert Evidenz für Unterschiede zwischen

- oral-akustischem und literat-schriftlichem Diskurs
- geskripteten und improvisierten oral-akustischen Texten
- interaktiven und passiven oral-akustischen Texten

Im Folgenden wird die unterschiedliche Verwendung von [Diskursmarkern](#)

in den genannten Textsorten nachgewiesen

- Diese intuitiv wahrgenommenen Unterschiede existieren tatsächlich
- Konnte Unterschiede zwischen oral-akustischem und literat-schriftlichem Diskurs nachweisen
- Des Weiteren auch Unterschiede zwischen geskripteten und improvisierten Texten
- und zwischen interaktiven Texten und eher passiven Texten
- basieren auf der Betrachtung von Diskursmarkern
- weil Diskursmarker einerseits von der relevanten Literatur in diesem Zusammenhang immer wieder genannt wurden und es andererseits aber bisher nicht besonders viel Forschung auf diesem Gebiet gibt

- oral-akustische Texte nutzen mehr Diskursmarker als literat-schriftliche
- improvisierte Texte nutzen mehr Diskursmarker als geskriptete
- interaktive Texte nutzen mehr Diskursmarker als passive

- Konkret:
- oral-akustische Texte nutzen mehr Diskursmarker als literat-schriftliche,
- improvisierte Texte mehr als geskriptete
- und interaktive mehr als passive

# BEGRIFF

## Diskursmarker

Wörter wie *and*, *but* und *so*

- keine inhaltliche Bedeutung
- signalisieren Beziehungen zwischen Diskurssegmenten
- Wegweiser im Text

Doch was sind überhaupt Diskursmarker?

- später genauer
- für den Anfang reicht:
  - keine inhaltliche Bedeutung
  - Wegweiser im Text
  - indem sie Beziehungen zwischen Diskurssegmenten anzeigen

## BEGRIFF Diskurs

Einheit von Sprache, länger als ein einzelner Satz

Quelle

5

## BEGRIFF Diskurs

Einheit von Sprache, länger als ein einzelner Satz

Quelle

Literat – Konzept für **schriftliches** Medium

- » literat-schriftliche Texte
- » Readability

Oral – Konzept für **akustisches** Medium

- » oral-akustische Texte
- » Listenability

5 | Johann Fischer, 4.10.20

5 | Johann Fischer, 4.10.20

## BEGRIFF Diskurs

Einheit von Sprache, länger als ein einzelner Satz

Quelle

Literat – Konzept für **schriftliches** Medium

- » literat-schriftliche Texte
- » Readability

Oral – Konzept für **akustisches** Medium

- » oral-akustische Texte
- » Listenability

Medien	Konzepte	
	literat	oral
schriftlich	Stummlesen eines Zeitungsartikels	Stummlesen einer Chatnachricht
akustisch	Ahören eines vorgelesenen Zeitungsartikels	Personelles Gespräch

5 | Johann Fischer, 4.10.20

- Diskurs: in der Linguistik: „Einheit von Sprache, die länger ist als ein einzelner Satz“
- ich unterscheide zwischen literat-schriftlichem und oral-akustischem Diskurs
- literat & oral: Konzepte, schriftlich & akustisch: Medien
- literat-schriftlich: Text konzipiert für das schriftliche Medium
- oral-akustisch: Text konzipiert für das akustische Medium
- auch andere Kombinationen:
  - Chatnachricht (oral-schriftlich)
  - vorgelesener Zeitungsartikel (Literat-akustisch)
- Verständlichkeit:
  - schriftliches Medium: Readability
  - akustisches Medium: Listenability

## MOTIVATION

- Verwendung von Sprachassistenten zum Vorlesen von z.B. Zeitungsartikeln
- Zeitungsartikel wurde geschrieben, um gelesen zu werden
- » vorgelesen ggf. nicht mehr so gut verständlich

## MOTIVATION

- Verwendung von Sprachassistenten zum Vorlesen von z.B. Zeitungsartikeln
- Zeitungsartikel wurde geschrieben, um gelesen zu werden
- » vorgelesen ggf. nicht mehr so gut verständlich

Iterates Konzept – akustisches Medium 

Iterates Konzept – schriftliches Medium 

6 | Information Literacy 2020

## MOTIVATION

- Verwendung von Sprachassistenten zum Vorlesen von z.B. Zeitungsartikeln
- Zeitungsartikel wurde geschrieben, um gelesen zu werden
- » vorgelesen ggf. nicht mehr so gut verständlich

Iterates Konzept – akustisches Medium 

Iterates Konzept – schriftliches Medium 

6 | Information Literacy 2020

## MOTIVATION

- Verwendung von Sprachassistenten zum Vorlesen von z.B. Zeitungsartikeln
- Zeitungsartikel wurde geschrieben, um gelesen zu werden
- » vorgelesen ggf. nicht mehr so gut verständlich

Iterates Konzept – akustisches Medium 

Iterates Konzept – schriftliches Medium 

Astronomers are searching the cosmic lost-and-found for use of the biggest, baddest black holes thought to exist. So far they have found a few dozen; it has become part of astronomical lore. If not quite a law, that at the center of every luminous city of light, called a galaxy, lurks something like a hungry black hole, a giant black hole that has devoured billions of stars. In some galaxies billions have disappeared. The bigger the galaxy, the more massive the black hole at its center.

By Dennis Overbye  
Published Jan. 15, 2011 Updated Jan. 20, 2011

6 | Information Literacy 2020

- Ursprung dieser Arbeit

- Am Anfang: was, wenn ein Sprachassisten Zeitungsartikel vorliest, für schriftliches Medium konzipiert
- kann sein, dass der Text auf diese Weise nicht mehr so gut verständlich ist
- **Beispiel** demonstrieren
- zuerst: Satz aus einem Zeitungsartikel an, v. Siri
- dann Satz selbst lesen + vergleichen, was besser verständlich war
- **zeigen**
- vielleicht war das Lesen jetzt auch einfacher, weil Sie den Satz zuvor schon gehört haben
- Verstehen beim Zuhören schwieriger als bei Lesen



- liegt daran, dass Listenability und Readability nicht das gleiche sind
- da Schrift und geschriebene Texte sich aus Sprache, aus Erzählungen entwickelt haben, ist ein Text mit hoher Listenability auch immer gut lesbar
- doch umgekehrt gilt das nicht
- obwohl Readability gut erforscht + viele Methoden, sie zu messen,
- lässt sich zu Listenability kaum Forschung finden
- auch kaum Methoden zur Messung
- für die linguistischen Features werden häufig Readability Skalen verwendet
- das scheint für Listenability jedoch nicht auszureichen

MOTIVATION

Wie können Texte so formuliert werden,  
dass sie über beide Medien funktionieren?

8 | Johannes Scherer, 4.1.2010

MOTIVATION

Wie können Texte so formuliert werden,  
dass sie über beide Medien funktionieren?

▼

Welche Faktoren erhöhen die **Listenability** eines Textes?

8 | Johannes Scherer, 4.1.2010

- Text soll über beide Medien funktionieren, ohne unbedingt das Konzept ändern zu müssen
  - > denn orales Konzept beinhaltet ggf. Redundanzen und Füllwörter
- Welche Faktoren erhöhen die Listenability eines Textes?

LISTENABILITY – FAKTOREN	
<b>Forschung</b>	<b>Ratgeber</b>
<ul style="list-style-type: none"> <li>• »Einfache« Texte [Flesch, Harwood, Chall &amp; Dial] • Einfache Wörter</li> <li>• Kurze Sätze [Ortmann &amp; Dipper] • Kurze Sätze</li> <li>• Wiederholungen [Ortmann &amp; Dipper] • Wiederholungen</li> <li>• Koordination (u.a. Diskursmarker) [Ortmann &amp; Dipper] • Zahlen runden • Bindewörter / Diskursmarker</li> </ul>	

© | johanneslechner.at/2016

- existierende Forschung:

- „einfachen“ Texte fast überall
- selten konkreter
- 2019, Ortmann und Dipper: orale Texte haben kürzere Sätze und mehr Wiederholungen als schriftliche
- außerdem mehr Koordination, bzw. Diskursmarker
- Journalisten und Schreiber fürs Radio oder Theater empfehlen zusätzlich: einfach Wörter, Runden von komplizierten Zahlen
- Kurze Sätze und Wiederholungen lassen sich einfach realisieren, doch was ist mit Diskursmarkern?
- Nutzen, um einen Flow im Text zu kreieren, wo im schriftlichen sonst z.B. Überschriften, Tabellen und Listen helfen
- Doch konkrete Regeln gibt es hier nicht
- die Satzlänge, die Schwierigkeit eines Wortes, die Anzahl an Wiederholungen lassen sich einfach messen, aber Diskursmarker sind nicht ganz so einfach zu verwenden
- und vor allem: Was sind eigentlich Diskursmarker?

## DISKURSMARKER Begriff

Nach Das et al., 2018:<sup>11</sup>  
(EnDimLex)

- lexikalischer Ausdruck, kann nicht flektiert werden
- signalisiert zweiseitige Relation zwischen Diskurssegmenten
- feststehender, nicht modifizierbarer Ausdruck
  - » nicht: *this reason* (for this *exact* reason)
- nicht semantisch kombinierbar
  - » nicht: *particularly if*
- feststehende Phrasen ok: *even if*

10 | Diskursmarker als Kriterium

## DISKURSMARKER Begriff

Nach Das et al., 2018:<sup>11</sup>  
(EnDimLex)

- lexikalischer Ausdruck, kann nicht flektiert werden
- signalisiert zweiseitige Relation zwischen Diskurssegmenten
- feststehender, nicht modifizierbarer Ausdruck
  - » nicht: *for this reason* (for this *exact* reason)
- nicht semantisch kombinierbar
  - » nicht: *particularly if*
- feststehende Phrasen ok: *even if*

10 | Diskursmarker als Kriterium

## DISKURSMARKER Begriff

Nach Das et al., 2018:<sup>11</sup>  
(EnDimLex)

- lexikalischer Ausdruck, kann nicht flektiert werden
- signalisiert zweiseitige Relation zwischen Diskurssegmenten
- feststehender, nicht modifizierbarer Ausdruck
  - » nicht: *for this reason* (for this *exact* reason)
- nicht semantisch kombinierbar
  - » nicht: *particularly if*
- feststehende Phrasen ok: *even if*

10 | Diskursmarker als Kriterium

- Begriff „Diskursmarker“ nicht eindeutig

- Verschiedene Begriffe:

- Cue Phrase

- Discourse Connective

- Connective

- Begriff nicht klar definiert:

- verschiedene Ansätze

- Die in der Computerlinguistik verwendete Definition ist aber gut zusammengefasst mit den Kriterien, die 2018 von den Autoren der EnDimLex Liste aufgestellt wurden

- kostenlose Liste Englischer Discourse Connectives

- Kriterien: s. Folie

- Beispiel:

- Satz ergibt zwar Sinn, aber ohne Kontext nicht eindeutig

- Diskursmarker zeigt Beziehung der Sätze (Vergleich, Kontrast)

-> Wegweiser im Text

Diskursmarker setzen sich aus verschiedenen anderen Wortgruppen zusammen  
 » erschwert automatische Erkennung beim String-Matching

u | Johanna Sieber; a.2010

Diskursmarker setzen sich aus verschiedenen anderen Wortgruppen zusammen  
 » erschwert automatische Erkennung beim String-Matching

Gollum was very angry as a result of Bilbo winning the ring. (*Adverb*)

u | Johanna Sieber; a.2010

Diskursmarker setzen sich aus verschiedenen anderen Wortgruppen zusammen  
 » erschwert automatische Erkennung beim String-Matching

Gollum was very angry as a result of Bilbo winning the ring. (*Adverb*)

Bilbo won the ring. As a result, Gollum was very angry. (*Diskursmarker*)

u | Johanna Sieber; a.2010

## -Diskursmarker mit Stringmatching

-Problem: sind funktionale Gruppe, die sich aus anderen Wortgruppen zusammensetzt

-> u.a. aus Adverbien, Konjunktionen, Präpositionalphrasen

-> d.h. für Erkennung z.B. Part-of-Speech Tagging nach dem Ausschlussverfahren

-> Beim 1. Satz würde ein Adverb erkannt werden

-> Beim 2. nicht. Dann müsste man die nicht identifizierten Worte und Phrasen noch mal genauer betrachten

-> Für unsere Zwecke reicht das Ergebnis aber aus, auch wenn Homographie eines Diskursmarkers mitgezählt werden

# DISKURSMARKER

## Zusammenfassung

- keine inhaltliche Bedeutung
- signalisieren Beziehungen zwischen Diskurssegmenten
- funktionale Gruppe
  - » setzt sich aus verschiedenen Wortgruppen zusammen

## DISKURSMARKER Bedeutungsgruppen

Nach Das et al., 2018 [d]  
(EnDimLex)

## DISKURSMARKER Bedeutungsgruppen

Nach Das et al., 2018 [ii]  
(EnDimLex)

**COMPARISON** Vergleich  
*but, although, in contrast, still, while, yet, ...*

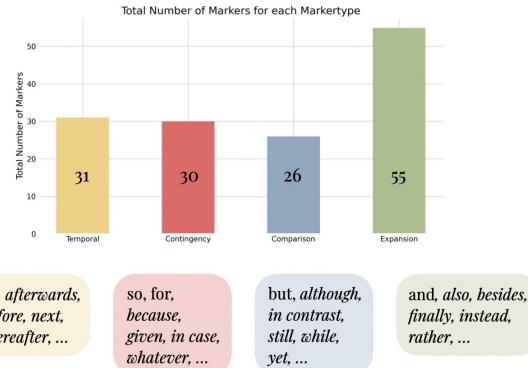
**CONTINGENCY** Folgern  
*so, for, because, given, in case, whatever, ...*

**EXPANSION** Hinzufügen eines Aspektes  
*and, also, besides, finally, instead, rather, ...*

**TEMPORAL** Zeitlicher Bezug  
*afterwards, as, before, next, thereafter, ...*

13 | Informativ/Sachlich, 4.5.2019

- DM sind funktionale Gruppe
- mehr ins Detail:
- welche Beziehung wird je genau von den einzelnen Markern signalisiert
- auch hier verschiedene Ansätze für die Einteilung in Bedeutungsgruppen
- Die meisten stimmen aber im Grunde miteinander überein, sie nennen die Gruppen nur unterschiedlich oder treffen detailliertere Unterscheidungen
- In unsere Betrachtung 4 Gruppen
  - Comparison Marker, die einen Vergleich zwischen den Segmenten anzeigen (but, although, in contrast)
  - Contingency Marker, die anzeigen, dass ein Segment sich aus dem anderen folgern lässt (so, for, given)
  - Expansion Marker, die signalisieren, dass ein Aspekt hinzugefügt wird (and, also, besides)
  - und Temporal Marker, die die Segmente zeitlich miteinander in Bezug setzen (afterwards, as, before, next)



- Die verwendete Liste hat folgende Verteilung auf die vier Klassen
- mit 55 der 149 Marker haben die Expansion Marker den weitaus größten Anteil
- zu dieser Gruppe gehören Wörter wie and, also, besides und finally
- die zweitgrößte Gruppe sind mit 31 Markern die Temporal Marker,
- dicht gefolgt von den 30 Contingency Markern
- und mit 26 Markern enthält die Comparison Klasse die wenigsten Marker

Diskursmarker können in mehreren Klassen gleichzeitig fallen:

Diskursmarker können in mehreren Klassen gleichzeitig fallen:

Sam and Pippin crouched behind a large tree-hole, **while** Frodo crept back a few yards towards the lane.  
(Temporal & Comparison)

Diskursmarker können in mehreren Klassen gleichzeitig fallen:

Sam and Pippin crouched behind a large tree-hole, **while** Frodo crept back a few yards towards the lane.  
(Temporal & Comparison)

**Since** they were all hobbits, and were trying to be silent, they made no noise that even hobbits would hear.  
(Contingency)

Diskursmarker können in mehreren Klassen gleichzeitig fallen:

Sam and Pippin crouched behind a large tree-hole, **while** Frodo crept back a few yards towards the lane.  
(Temporal & Comparison)

**Since** they were all hobbits, and were trying to be silent, they made no noise that even hobbits would hear.  
(Contingency)

I came also upon two others, but they turned away southward. **Since** then I have searched for your trail.  
(Contingency)

- Problem: DM kann je nach Kontext in verschiedene Klassen fallen kann, oder sogar in mehrere gleichzeitig
- Im ersten Satz: Zeitliche und Vergleich
- im 2. wird Since als Begründung verwendet
- im 3. jedoch im zeitlichen Sinne. Hier ist außerdem nicht 100% klar, ob Since (then) überhaupt ein DM ist oder nicht.
- Auch dieses Problem war für unsere Zwecke jedoch nicht so schlimm

## TEXTDATEN

### Corpora

akustische Corpora mit Transkripten von Audiomaterial &  
schriftliche Corpora mit ursprünglich schriftlichem Material

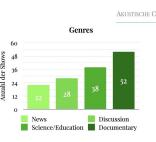
KRITERIEN	kostenlos groß qualitativ hochwertig nachrichtenähnlich
-----------	--

38 | Johannes Scherzer, Universität Regensburg

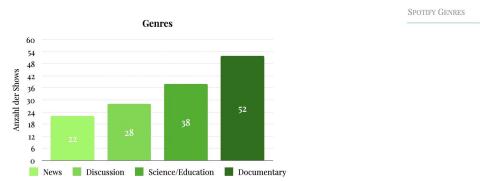
- um Vorkommen von Diskursmarkern zu untersuchen: Texte
- da Vergleich zwischen schriftlichem und akustischem Diskurs: schriftliche und akustische Texte
- wobei akustische Texte Transkripte von Audiomaterial sind, und schriftliche eben Texte, die ursprünglich schriftlich sind.
- schriftliche Corpora bereits vorhanden, akustische Corpora erstmal finden
- wichtig: kostenlos, groß, gute Qualität, nachrichtenähnlich (Conversational News)

CORPUS	Spotify Podcast Corpus	TED-LIUM 3 Corpus
DATEN	<ul style="list-style-type: none"><li>• fast 60.000 Stunden transkribiertes Audiomaterial</li><li>• verschiedene Produzenten</li><li>• WER: 18,1 %</li></ul>	<ul style="list-style-type: none"><li>• 1.983 TED-Talks</li><li>• ca. 4 Mio. Wörter</li><li>• WER: 6,7 %</li></ul>

- letztendlich zwei akustische Corpora: Spotify Podcast Corpus und TED-LIUM 3 Corpus
- Spotify Corpus: 100.000 Spotify Podcasts mit 60.000 h transkribiertes Audiomaterial, WER 18.1%
- TED Corpus: knapp 2000 TED-Talks mit ca. 4 Mio Wörtern
- Er hat eine WER von 6,7 %

AKUSTISCHE CORPORA – SPOTIFY																			
CORPUS	Spotify Podcast Corpus																		
DATEN	<ul style="list-style-type: none"> <li>fast 60.000 Stunden transkribiertes Audiomaterial</li> <li>verschiedenste Produzenten</li> <li>WER: 18,1 %</li> </ul>																		
RELEVANT	<ul style="list-style-type: none"> <li>140 Shows, 2.782 Episoden</li> <li>ca. 17 Mio. Wörter</li> </ul>																		
 <p>Genres</p> <table border="1"> <thead> <tr> <th>Genre</th> <th>Anzahl der Shows</th> </tr> </thead> <tbody> <tr> <td>News</td> <td>22</td> </tr> <tr> <td>Science-Education</td> <td>28</td> </tr> <tr> <td>Discussion</td> <td>38</td> </tr> <tr> <td>Documentary</td> <td>53</td> </tr> </tbody> </table>  <p>Konversationsarten</p> <table border="1"> <thead> <tr> <th>Konversationsart</th> <th>Anzahl der Shows</th> </tr> </thead> <tbody> <tr> <td>Monolog</td> <td>16</td> </tr> <tr> <td>Kooperativer Monolog</td> <td>38</td> </tr> <tr> <td>Dialog</td> <td>66</td> </tr> </tbody> </table>		Genre	Anzahl der Shows	News	22	Science-Education	28	Discussion	38	Documentary	53	Konversationsart	Anzahl der Shows	Monolog	16	Kooperativer Monolog	38	Dialog	66
Genre	Anzahl der Shows																		
News	22																		
Science-Education	28																		
Discussion	38																		
Documentary	53																		
Konversationsart	Anzahl der Shows																		
Monolog	16																		
Kooperativer Monolog	38																		
Dialog	66																		

- Spotify: viele verschiedene Podcasts, nicht alle relevant
  - > nach News Shows durchsucht, zu wenige
- Dann nach sachliche Genres
  - > Documentary, History, Science und True Crime
- 140 Shows, knapp 2.800 Episoden, ca. 17 Mio. Wörter
- In 4 Genres und 3 Konversationsarten unterteilt



**NEWS** Fokus auf Nachrichten

**DISCUSSION** Fokus auf Diskussionen und Meinungsaustausch

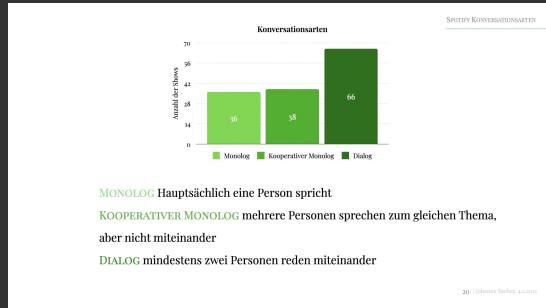
**SCIENCE/EDUCATION** übermitteln Wissen

**DOCUMENTARY** geskriptet, gut recherchiert, zu einem bestimmten Thema

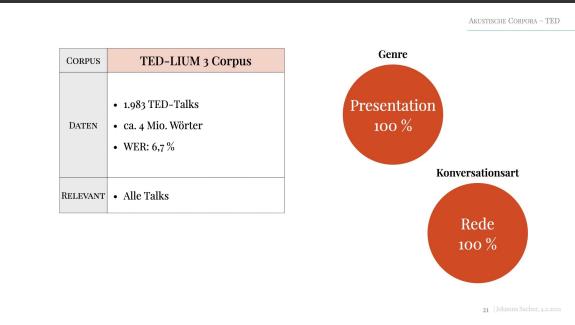
19 | Johanna Sicher, 4.2.2021

Genres: News, Discussion, Science/Education und Documentary

- News: Fokus auf Nachrichten, egal zu welchem Thema (22 Shows)
- Discussion: Fokus auf Diskussionen und Meinungsaustausch (28)
- Science/Education: Wissenschaftsshows oder Shows über sehr spezielle Themen, z.B. Podcasts von Fachleuten für andere Fachleute (38 Shows)
- Documentary: geskriptete, gut recherchierte Shows zu einem bestimmten Thema. Oft z.B. True Crime (52 Shows)



- Konversationsarten: Monolog, kooperativen Monolog und Dialog
- Monolog: hauptsächlich eine Person spricht (36 Shows)
- kooperativer Monolog: mehrere Personen sprechen zum gleichen Thema, aber nicht miteinander (38 Shows)
  - > Dokumentationen
- Dialog: zwei oder mehr Personen reden miteinander (66 Shows)



- TED: keine Infos über Thema des Talks, nur Redner und Titel
- > eigenes Genre und eine eigene Konversationsart für Talks
- TED Talks sind generell sehr speziell und haben ihren eigenen „Stil“



**PRESENTATION** vor einer Menge Zuhörer nach einem  
vorbereiteten Skript präsentiert

TED – GENRE

22 | Johanna Sacher, 4.3.2021

Konversationsart



**REDE** vor einer Menge Zuhörer nach einem  
vorbereiteten Skript gehalten

TED – KONVERSATIONSArt

23 | Johanna Sacher, 4.3.2021

SCHRIFFTLCHE CORPORA		
CORPUS	New York Times Corpus	
DATEN	<ul style="list-style-type: none"> <li>• 1,8 Mio. Nachrichtenartikel der New York Times</li> <li>• ca. 1,1 Mrd. Wörter</li> </ul>	
RELEVANT	<ul style="list-style-type: none"> <li>• Alles</li> </ul>	

24 | Johann Sohler, 4.2.2019

SCHRIFFTLCHE CORPORA		
CORPUS	New York Times Corpus	Gigaword Corpus
DATEN	<ul style="list-style-type: none"> <li>• 1,8 Mio. Nachrichtenartikel der New York Times</li> <li>• ca. 1,1 Mrd. Wörter</li> </ul>	<ul style="list-style-type: none"> <li>• Newswire Textdaten</li> <li>• aus 7 Quellen</li> <li>• ca. 4 Mrd. Wörter</li> </ul>
RELEVANT	<ul style="list-style-type: none"> <li>• Alles</li> </ul>	<ul style="list-style-type: none"> <li>• Alles</li> </ul>

24 | Johann Sohler, 4.2.2019

- schriftliche Corpora einfacher:
- New York Times Corpus
  - 1,8 Mio. Artikel, ca. 1.1 Mrd. Wörter
- außerdem Gigaword Corpus
  - ca. 4 Mrd. Wörter
- beiden nicht nach Genres und Konversationsarten aufgeteilt,  
-> enthalten nur schriftliche Nachrichtenartikel

TYP	Akustische Corpora		Schriftliche Corpora		
	CORPUS	Spotify	TED-LIUM 3 Corpus	New York Times	Gigaword
RELEVANT		<ul style="list-style-type: none"><li>• 2.782 Episoden</li><li>• ca. 17 Mio. Wörter</li></ul>	<ul style="list-style-type: none"><li>• 1.983 TED-Talks</li><li>• ca. 4 Mio. Wörter</li></ul>	<ul style="list-style-type: none"><li>• 1,8 Mio. Artikel</li><li>• 1,1 Mrd. Wörter</li></ul>	<ul style="list-style-type: none"><li>• ca. 4 Mrd. Wörter</li></ul>

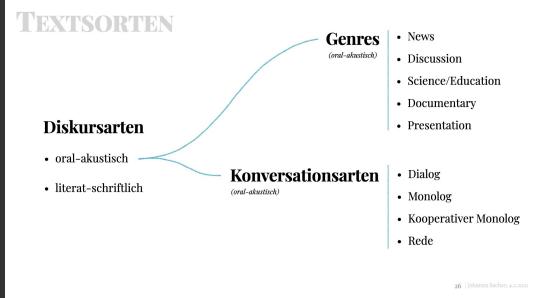
Wörter

## TEXTSORTEN

### Diskursarten

- oral-akustisch
- literat-schriftlich

16 | Johanna Fischer, Universität Regensburg



-verschiedene Textsorten:

-oral-akustische und literat-schriftliche

-oral-akustischen: Spotify, TED

-literat-schriftliche: NYT, Gigaword

-oral-akustische auf zwei Arten aufgeteilt:

-nach Genres

-und nach Konversationsarten

-gibt außerdem interaktivere Texte (Dialog) und passivere (Monolog, kooperativer Monolog)

-gibt sachliche Genres (News, Science, Documentary) und interaktionslastigere (Discussion, Presentation, Rede) bei denen mit Diskurspartnern oder den Zuhörern interagiert wird

# FRAGEN

1. Welche Textsorten stützen sich besonders auf Diskursmarker?
2. An welchen Positionen im Text stützen sich die jeweiligen Textsorten besonders auf Diskursmarker?
3. An welchen Positionen im Satz stützen sich die jeweiligen Textsorten besonders auf Diskursmarker?
4. Auf welche Klassen von Diskursmarkern stützen sich die jeweiligen Textsorten besonders?
5. Welche Diskursmarker werden innerhalb der jeweiligen Klassen besonders genutzt?

# AUSWERTUNG

## 1. Generelle Verteilung

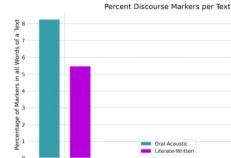
- oral-akustische Texte nutzen mehr Diskursmarker als literat-schriftliche
- improvisierte oral-akustische Texte nutzen mehr Diskursmarker als geskriptete

28 | Johanna Sacher, 4.1.2019

# AUSWERTUNG

## 1. Generelle Verteilung

- oral-akustische Texte nutzen mehr Diskursmarker als literat-schriftliche
- improvisierte oral-akustische Texte nutzen mehr Diskursmarker als geskriptete



P-Wert < 0.001

» oral-akustische Texte nutzen mehr Diskursmarker als literat-schriftliche

28 | Johanna Sacher, 4.1.2019

## 1.: Generellen Verteilung

-OA Texte nutzen mehr DM als LS

-Diagramm: durchschnittlicher prozentualer Anteil von Diskursmarkern an allen Wörtern der jeweiligen Texte,

-türkis abgebildet sind die OA Texte, pink die LS Texte

-deutlich: OA Texte (ca. 8%) mehr DM als LS (ca. 5%)

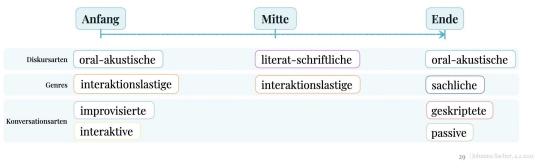
-Unterschied ist, wie alle hier vorgestellten Ergebnisse, statistisch signifikant

- außerdem: improvisierte OA Texte, (Diskussionen) nutzen mehr DM als geskriptete (Dokumentationen)

# AUSWERTUNG

## 2. Textpositionen

Texte, die an der jeweiligen Position im Text je **mehr Diskursmarker** nutzen als andere



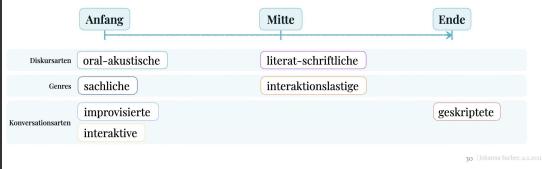
## 2: Position von Diskursmarkern im Text

- DM in allen Texten sehr gleichmäßig über den gesamten Text verteilt
- im Vergleich: oral-akustische Texte am Textanfang und Ende mehr DM nutzen als LS Texte.
- LS Texte nutzen dafür in der Mitte mehr DM als OA
- außerdem: interaktionslastige, interaktive und improvisierte Texte am Anfang mehr DM als andere,
- sachliche, geskriptete und passive Texte eher am Ende

# AUSWERTUNG

## 3. Satzpositionen

Texte, die an der jeweiligen Position im Satz je **mehr Diskursmarker** nutzen als andere

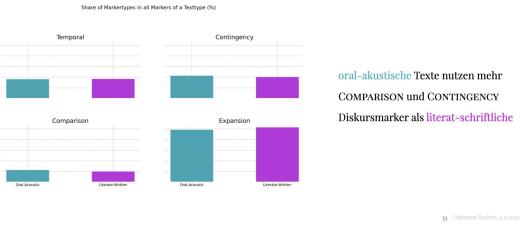


### 3: Position der DM im Satz

- innerhalb der Sätze: DM häufiger am Anfang des Satzes verwendet als am Ende
- Vergleich: OA am Anfang des Satzes mehr DM als LS
- LS Texte dafür in der Mitte des Satzes mehr DM als OA
- sachlichere Genres beginnen ihre Sätze eher mit DM als interaktionslastigere
- und improvisierte und interaktive Konversationsarten wie Dialoge beginnen ihre Sätze eher mit Diskursmarkern als geskriptete, passive, wie Texte des kooperativen Monologs

## AUSWERTUNG

### 4. Diskursmarker-Klassen



- Vergleich der Nutzung der vier Diskursmarker Klassen
- Temporal und Expansion Marker: von OA und LS Texten etwa gleich häufig
- OA: mehr Comparison und Contingency als LS
- > einfache, schriftliche Nachrichten sind überlegt formuliert
- > improvisierte Texte oder Texte die ihren Zuhörern was erklären wollen, mehr Vergleiche und Schlussfolgerungen anstellen

## AUSWERTUNG

### 5. Häufigste Diskursmarker der Klassen

Diskursarten

as TEMPORAL  
and EXPANSION  
but COMPARISON  
for CONTINGENCY

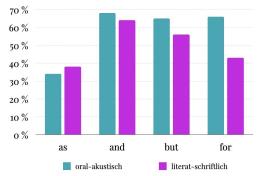
32 | Jürgen Schleher, a. 2009

## AUSWERTUNG

### 5. Häufigste Diskursmarker der Klassen

as TEMPORAL  
and EXPANSION  
but COMPARISON  
for CONTINGENCY

Diskursarten  
prozentualer Anteil der Marker an den Markern ihrer jeweiligen Klasse



32 | Jürgen Schleher, a. 2009

5: einzelnen Marker, bzw. welche Marker in den jeweiligen Klassen die häufigsten sind

-die häufigsten Marker für nahezu alle Textsorten sind dieselben

-nennen

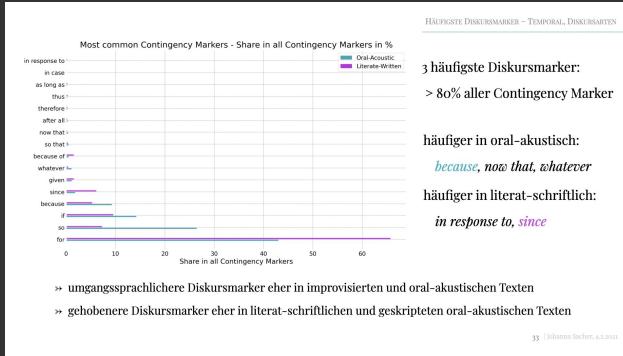
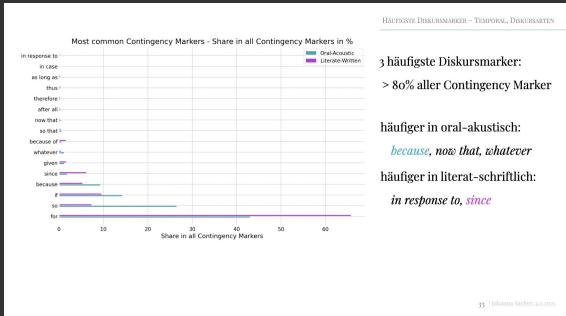
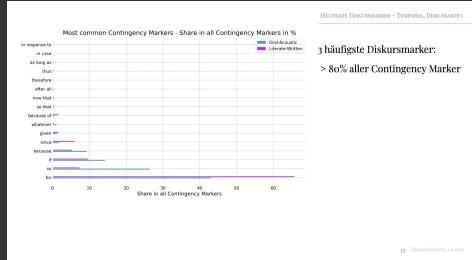
-Diagramm: prozentualer Anteil dieser Marker an allen Markern ihrer Klasse in den Diskursarten

-z.B. „and“: bei OA Anteil 68% an allen Expansion Markern

-bei LA 64%

-auch „but“ und „for“ machen bei OA jeweils mehr als die Hälfte aller Diskursmarker ihrer Klasse aus

-Es ist aber auch zu beachten, dass sie wahrscheinlich nicht an jeder Stelle als Diskursmarker benutzt wurden, das bleibt für kommende Arbeiten herauszufinden



## Diagramm: häufigste Contingency Marker für OA & LS

- die 3 häufigsten, for, so und if, machen insgesamt mehr als 80% der verwendeten Contingency Marker aus
- außerdem: Unterschied in den verwendeten Markern zw. Diskursarten
- because, now that und whatever: OA
- in response to und since: LS
- > umgangssprachlichere Diskursmarker wie „because“ und „whatever“ häufiger in OA - und dort auch häufiger in improvisierten Texten
- > gehobenere „since“ oder z.B. auch „although“ und „nonetheless“ häufiger in LS, oder auch in geskripteten OA

# ZUSAMMENFASSUNG

## AUSBlick

- Aussortieren von Homographen
- Akustischen Nachrichten-Corpus verwenden
- Listenability aller betrachteten Texte messen und mit Nutzung von Diskursmarkern in Zusammenhang setzen

- Zu tun: Aussortieren von Mehrdeutigkeiten
  - > z.B. Part-of-Speech Tagging - ob Wort DM oder nicht
- akustischen Corpus verwenden, der richtige Nachrichtensendungen enthält (Radionachrichten, Podcasts)
- Listenability mit geeigneter Methode messen, Ergebnisse mit der Nutzung von Diskursmarkern in Zusammenhang zu stellen
- sind davon ausgegangen, dass Listenability zum einen durch das orale Konzept von improvisierten, frei gesprochenen Podcasts gegeben ist, und dass zum anderen die Journalisten, die geskriptete Dokumentationen erstellen und die Redner der TED-Talks etwas von ihrem Handwerk verstehen
- es wäre sehr spannend, das mit tatsächlichen Zahlen belegen zu können

# ZUSAMMENFASSUNG

## AUSBlick

- Aussortieren von Homographen
- Akustischen Nachrichten-Corpus verwenden
- Listenability aller betrachteten Texte messen und mit Nutzung von Diskursmarkern in Zusammenhang setzen

## WICHTIG

Diskursmarker werden in verschiedenen Textsorten unterschiedlich eingesetzt

- » Auswirkung auf Listenability
- oral-akustische Texte nutzen mehr Diskursmarker als literat.-schriftliche
- improvisierte Texte nutzen mehr Diskursmarker als geskriptete
- interaktive Texte nutzen mehr Diskursmarker als passive

- DM in verschiedenen Textsorten unterschiedlich eingesetzt
- > Auswirkungen auf Listenability
- OA durchschnittlich mehr DM als LS
- improvisierte Texte mehr als geskriptete
- interaktivere Texte mehr als passive