

Diskursmarker

in
schriftlichem & akustischem
Diskurs

BACHELORVERTEIDIGUNG

Johanna Sacher, 4.2.2021



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

Außerdem:

- Was sind Diskursmarker
- Problematiken
- Bedeutungsgruppen

BEGRIFF

Diskurs

Einheit von Sprache, länger als ein einzelner Satz

Quelle

- Damit wir alle auf einem Stand sind,
- Zuerst Begriffe klären
 - Diskurs: Eine Einheit von Sprache, länger als ein einzelner Satz (Quelle?)

BEGRIFF

Diskursmarker

Wörter wie *and*, *but* und *so*

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

Diskursmarker:

- später genauer
- für den Anfang reicht es zu wissen, dass DM Wörter sind, die
 - keine inhaltliche Bedeutung zum Satz beitragen, sondern
 - eine Beziehung zw. zwei Diskurssegmenten signalisieren
 - und dadurch als Wegweiser im Text dienen

DISKURS

Literat vs. Oral

Literat – Konzept für das **schriftliche** Medium

- » literat-schriftliche Texte (LS)
- » Readability

Oral – Konzept für das **akustische** Medium

- » oral-akustische Texte (OA)
- » Listenability

- Literater Diskurs: Diskurs, der für das schriftliche Medium konzipiert wurde, als zum (stummen) Lesen gedacht ist
 - Wie gut Diskurs verständlich ist, der über das schriftliche Medium kommuniziert wird, wird im Englischen auch „Readability“ genannt
- Oraler Diskurs: Diskurs, der für das akustische Medium konzipiert wurde, also zum Anhören gedacht ist
 - Wie gut Diskurs verständlich ist, der über das akustische Medium kommuniziert wird, wird auch „Listenability“ genannt

Medien	Konzepte	
	literat	oral
schriftlich	Stummes lesen eines Zeitungsartikels	Stummes lesen eines YouTube Kommentars
akustisch	Anhören eines vorgelesenen Zeitungsartikels	Persönliches Gespräch

- Medien und Konzepte können beliebig kombiniert werden
- jedoch sollte dabei nie die Verständlichkeit vergessen werden,
- was uns zu der eigentliche Motivation dieser Arbeit bringt

MOTIVATION

Wieso dieses Thema?

Verwendung von Sprachassistenten zum Vorlesen von z.B. Zeitungsartikeln



8 | Johanna Sacher, 4.2.2021

- Sprachassistenten können Dinge vorlesen, die wir zuvor selbst lesen mussten, z.B. Zeitungsartikel
- gibt uns als Nutzern freie Hände, um parallel ggf. andere Aufgaben zu erledigen oder einfach zu entspannen
- Beispiele zeigen / vorspielen

Zeitungsartikel wurde geschrieben, um gelesen zu werden

» vorgelesen ggf. nicht mehr so gut verständlich

- Zeitungsartikel wurden geschrieben, um gelesen zu werden
 - sind über das neue, akustische Medium ggf. nicht mehr so gut verständlich

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

→ Wie können Zeitungsartikel oder Texte allgemein (auf einfache Weise) so formuliert werden, dass sie über beide Medien funktionieren?

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



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

- Kurze Sätze
- Einfache Wörter
- Zahlen runden
- Koordination / Bindewörter / Diskursmarker

- gibt kaum Forschung zu Listenability

DISKURSMARKER

Begriff

- Begriff ist nicht eindeutig definiert
- Verschiedene Begriffe in Benutzung

12 | Johanna Sacher, 4.2.2021

- Begriff ist nicht eindeutig definiert
- Verschiedene Begriffe:
 - Cue Phrase
 - Discourse Connective
 - Connective
- v.a. wegen verschiedener Forschungsansätze

Funktionale Definition nach Bruce Fraser

- DM ist ein lexikaler Ausdruck
- In <S1 S2> muss ein DM Teil von S2 sein
- DM trägt nicht zur semantischen Bedeutung der Sequenz bei, sondern signalisiert eine Relation zwischen S1 und S2

- Bruce Fraser ist der Forscher mit den meisten veröffentlichten Arbeiten über Diskursmarker, er beschäftigt sich bereits seit über 30 Jahren mit ihnen
- Er plädiert dafür, Diskursmarker offiziell als funktionale Wortgruppe anzuerkennen und definiert sie ausführlich
- Fraser geht davon aus, dass jede Sprache eine funktionale Klasse lexikalischer Ausdrücke hat, die er „pragmatic markers“ nennt
- diese pragmatic markers sind Teil des Diskurs Segments, aber nicht unbedingt Teil der Bedeutung
- sie signalisieren vielmehr Aspekte, die der Sprecher nicht explizit ausspricht
- Diskursmarker sind in dieser Betrachtung eine Untergruppe der pragmatic markers
- sie signalisieren eine Beziehung zwischen dem beinhaltenden Diskurssegment, meistens S2, und einem vorangegangenen Segment S1

I love the Shire. But I begin to wish, somehow, that I had gone, too.

- lässt man das But weg, hat der Satz noch dieselbe Bedeutung, er ist aber, vor allem ohne Kontext, schwieriger zu verstehen.
- wohin gegangen?
- Auf den ersten Blick nur And, doch eigentlich ist es eine DM-Phrase: „And also“ und eine weitere Phrase, die Fraser als commentary pragmatic marker einordnen würde: „I fancy“.
- „And also“ signalisiert, dass das beinhalten Segment das vorherige weiter ausführt, bzw. etwas hinzufügt.
- Beide Phrasen können weggelassen werden, ohne dass sich die Bedeutung des Satzes ändert, doch ist er dann nicht mehr so einfach verständlich wie zuvor

I love the Shire. But I begin to wish, somehow, that I had gone, too.

I love the Shire. I begin to wish, somehow, that I had gone, too.

I love the Shire. But I begin to wish, somehow, that I had gone, too.

I love the Shire. I begin to wish, somehow, that I had gone, too.

You are the master of Bag End now. And also, I fancy, you'll find a golden ring.

I love the Shire. But I begin to wish, somehow, that I had gone, too.

I love the Shire. I begin to wish, somehow, that I had gone, too.

You are the master of Bag End now. And also, I fancy, you'll find a golden ring.

You are the master of Bag End now. You'll find a golden ring.

Kriterien des EnDimLex

- DM ist ein lexikaler Ausdruck und kann nicht flektiert werden
- DM signalisiert eine zweiseitige Relation, deren Argumente abstrakte Objekte sind
- Argumente können in Klausel-, Satz- oder Phrasenstruktur ausgedrückt werden

- Fraser gibt keine vollständige Liste von DM an, höchstens Beispiele
- eine kostenlose Liste englischer DiscourseConnectives ist die EnDimLex Liste
- sie enthält 149 Einträge
- die Kriterien ähneln der von Fraser aufgestellten Definition stark

Weitere Bedingungen

- feststehender, nicht modifizierbarer Ausdruck
 - » nicht: *for this reason* (*for this **exact** reason*)
- darf nicht semantisch kombinierbar sein
 - » nicht: *particularly if*
 - » feststehende Phrasen sind ok: *even if*

- für die Aufnahme in die Liste wurden außerdem ergänzende Kriterien aufgestellt
- zum einen muss ein potentieller Ausdruck feststehend und nicht modifizierbar sein
 - > „for this reason“ wird nicht aufgenommen, da zu „for this exact reason“ modifizierbar
- außerdem sind keine Ausdrücke erlaubt, die semantisch kombinierbar sind
 - > also z.B. Phrasen wie „particularly if“
 - > feststehende Phrasen wie „even if“ hingegen wurden aufgenommen

DISKURSMARKER

Vergleich

Funktionale Definition (Fraser)		EnDimLex-Kriterien
lexikaler Ausdruck	✓	lexikaler Ausdruck
	→	kann nicht flektiert werden
signalisiert Relation zwischen Diskurssegmenten	✓	signalisiert zweiseitige Relation zwischen Klauseln, Sätzen oder Phrasen
trägt nicht zur Bedeutung des Satzes bei	←	
meistens Adverbien, Konjunktionen, Präpositionalphrasen	✓	meistens Adverbien, Konjunktionen, Präpositionalphrasen

DM setzen sich aus verschiedenen anderen Wortgruppen zusammen
» erschwert automatische Erkennung

DM setzen sich aus verschiedenen anderen Wortgruppen zusammen

» erschwert automatische Erkennung

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

DM setzen sich aus verschiedenen anderen Wortgruppen zusammen

» erschwert automatische Erkennung

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

DM setzen sich aus verschiedenen anderen Wortgruppen zusammen

» erschwert automatische Erkennung

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

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

DM setzen sich aus verschiedenen anderen Wortgruppen zusammen

» erschwert automatische Erkennung

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

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

DISKURSMARKER

Zusammenfassung

- keine inhaltliche Bedeutung
- signalisieren Beziehungen zwischen Diskurssegmenten
- setzen sich aus verschiedenen Wortgruppen zusammen
- automatische Erkennung schwierig

DISKURSMARKER

Bedeutungsgruppen

Gibt verschiedene Ansätze, DM anhand ihrer Bedeutung in Klassen aufzuteilen

Fraser

CONTRASTIVE MARKERS Kontrast zwischen S1 und S2

but, *alternatively*, *although*, *even so*, *still*, *yet*, ...

Fraser

CONTRASTIVE MARKERS Kontrast zwischen S1 und S2

but, *alternatively, although, even so, still, yet, ...*

ELABORATIVE MARKERS Genauere Ausführung von S1 in S2

and, *also, besides, for instance, moreover, similarly, ...*

Fraser

CONTRASTIVE MARKERS Kontrast zwischen S1 und S2

but, *alternatively, although, even so, still, yet, ...*

ELABORATIVE MARKERS Genauere Ausführung von S1 in S2

and, *also, besides, for instance, moreover, similarly, ...*

INFERENTIAL MARKERS S2 kann aus S1 gefolgert werden

so, *consequently, therefore, thus, ...*

EnDimLex

COMPARISON Vergleich

but, although, in contrast, still, while, yet, ...

EnDimLex

COMPARISON Vergleich

but, although, in contrast, still, while, yet, ...

CONTINGENCY Folgern, Möglichkeiten aufzeigen

so, for, because, given, in case, whatever, ...

EnDimLex

COMPARISON Vergleich

but, although, in contrast, still, while, yet, ...

CONTINGENCY Folgern, Möglichkeiten aufzeigen

so, for, because, given, in case, whatever, ...

EXPANSION Hinzufügen eines Aspektes

and, also, besides, finally, instead, rather, ...

EnDimLex

COMPARISON Vergleich

but, although, in contrast, still, while, yet, ...

CONTINGENCY Folgern, Möglichkeiten aufzeigen

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, ...

EnDimLex	Fraser	Funktion
Comparison	Contrastive	Vergleich, Kontrast
Contingency	Inferential	Folgern
Expansion	Elaborative	Ausführen, Illustrieren
Temporal	Elaborative, Inferential	Zeitlich in Bezug setzen

DM können in mehrer Klassen gleichzeitig fallen:

DM können in mehrer Klassen gleichzeitig fallen:

Sam and Pippin crouched behind a large tree-bole, **while** Frodo crept back a few yards towards the lane.

(Temporal & Comparison)

DM können in mehrer Klassen gleichzeitig fallen:

Sam and Pippin crouched behind a large tree-bole, **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)

DM können in mehrer Klassen gleichzeitig fallen:

Sam and Pippin crouched behind a large tree-bole, **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.

(Temporal)

TEXTSORTEN

Diskursarten

- oral-akustisch
- literat-schriftlich

TEXTSORTEN

Genres

- News
- Discussion
- Science/Education
- Documentary
- Presentation

TEXTSORTEN

Konversationsarten

- Dialog
- Monolog
- Kooperativer Monolog
- Rede

TEXTDATEN

Corpora

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

- kostenlos
- groß
- qualitativ hochwertig
- nachrichtenähnlich

- Anforderungen:

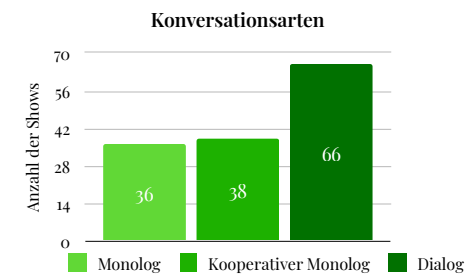
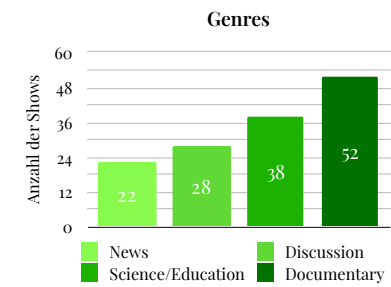
- kostenlos
- groß
- qualitativ hochwertig
- nachrichtenähnlich

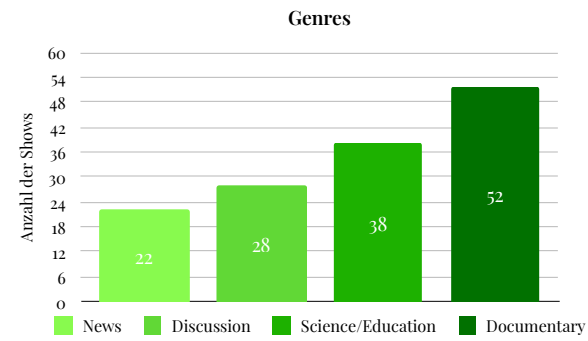
- kurz Probleme für akustische Corpora nennen:

- fehlende Interpunktion
- WER

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

CORPUS	Spotify Podcast Corpus
DATEN	<ul style="list-style-type: none">• fast 60.000 Stunden transkribiertes Audiomaterial• verschiedenste Produzenten• WER: 18,1 %
NUTZBAR	<ul style="list-style-type: none">• 140 Shows, 2.782 Episoden• ca. 17 Mio. Wörter



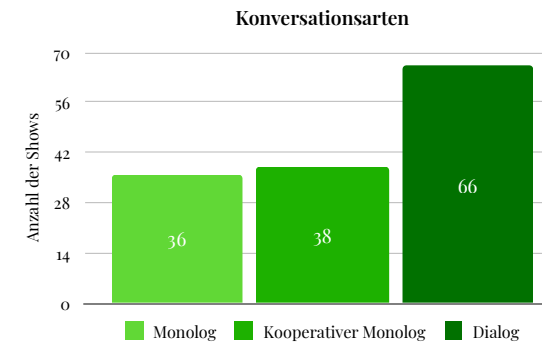


NEWS Fokus auf Nachrichten

DISCUSSION Fokus auf Diskussionen und Meinungs Austausch

SCIENCE/EDUCATION übermitteln Wissen

DOCUMENTARY geskriptet, gut recherchiert, zu einem bestimmten Thema



MONOLOG Hauptsächlich eine Person spricht

KOOPERATIVER MONOLOG mehrere Personen sprechen zum gleichen Thema, aber nicht miteinander

DIALOG mindestens zwei Personen reden miteinander

CORPUS	TED-LIUM 3 Corpus
DATEN	<ul style="list-style-type: none">• 1.983 TED-Talks• ca. 4 Mio. Wörter• WER: 6,7 %
NUTZBAR	<ul style="list-style-type: none">• Alle Talks



Genre



PRESENTATION vor einer Menge Zuhörer nach einem
vorbereiteten Skript präsentiert

Konversationsart



REDE vor einer Menge Zuhörer nach einem
vorbereiteten Skript gehalten

CORPUS	Spotify Podcast Corpus	TED-LIUM 3 Corpus
DATEN	<ul style="list-style-type: none">• fast 60.000 Stunden transkribiertes Audiomaterial• verschiedenste Produzenten• WER: 18,1 %	<ul style="list-style-type: none">• 1.983 TED-Talks• ca. 4 Mio. Wörter• WER: 6,7 %
NUTZBAR	<ul style="list-style-type: none">• 140 Shows, 2.782 Episoden• ca. 17 Mio. Wörter	<ul style="list-style-type: none">• Alle Talks

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

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

TYP	Akustische Corpora		Schriftliche Corpora	
CORPUS	Spotify	TED-LIUM 3 Corpus	New York Times	Gigaword
NUTZBAR	<ul style="list-style-type: none">• 140 Shows, 2.782 Episoden• ca. 17 Mio. Wörter	<ul style="list-style-type: none">• 1.983 TED-Talks• ca. 4 Mio. Wörter	<ul style="list-style-type: none">• 1,8 Mio. Artikel• 1,1 Mrd. Wörter	<ul style="list-style-type: none">• ca. 4 Mrd. Wörter

VORGEHEN

Diskursmarker im Text erkennen

Diskursmarker wurden mit Hilfe eines einfachen
String-Matching Verfahren mit den Texten gematched

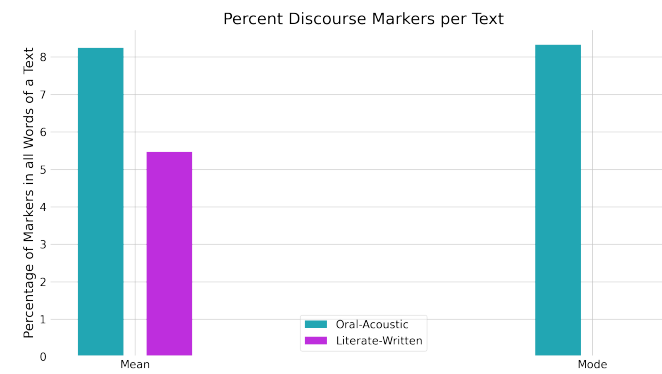
FRAGEN

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 mehr als literat-schriftliche
- improvisierte OA Texte mehr als geskriptete



P-Wert < 0.001

» OA nutzt mit einer Effektgröße (EG)
von 1,63 mehr DM als LS

AUSWERTUNG

2. Textpositionen

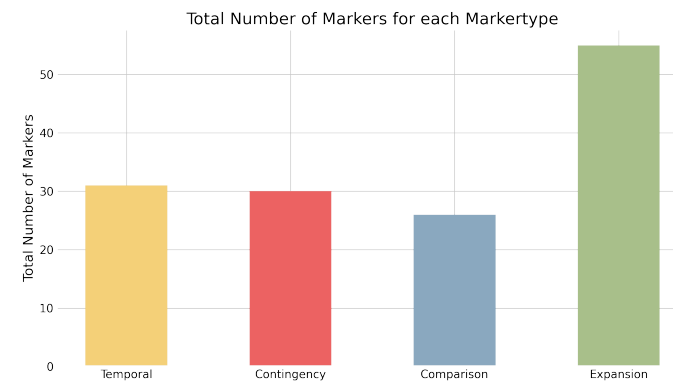
- OA mehr am Anfang und am Ende, LS mehr in der Mitte
- interaktionslastigere Texte mehr am Anfang und in der Mitte, sachlichere mehr am Ende
- improvisierte mehr am Anfang, gespkriptete mehr am Ende

AUSWERTUNG

3. Satzpositionen

- OA mehr am Anfang, LS mehr in der Mitte
- sachliche Genres (News, Science) mehr am Anfang und weniger in der Mitte
- interaktionslastigere (Dialog) mehr am Anfang und weniger am Satzende, passive (Koop. Monolog) mehr am Satzende

4. Diskursmarker-Klassen



größter Anteil an allen DM in
allen Texten: EXPANSION DM
(u.a. *and*)

AUSWERTUNG

5. Häufigste Diskursmarker

AUSBlick

Offene Fragen

Was bleibt zu tun?

ZUSAMMENFASSUNG

Überblick über die wichtigsten Ergebnisse

