

Deutsche Adverbien und Adverbiale aus der Sicht des TreeTaggers

Céline Largier Vié

Sorbonne Nouvelle Université des Cultures (CEREG – EA 4223)

Eva Schaeffer-Lacroix

Sorbonne Université | Inspé Paris) (CeLiSo – EA 7332)

Inhalt

- Unser Datensatz
- Audiodeskription
- TreeTagger
- Definitionen im Stuttgart-Tübingen-Tagset 1.0.
- Adverbien und Adverbiale in unserem Datensatz
- Fehlerbestand und –analyse
- Menschliche versus maschinelle Analyse der Form "etwas"
- Offene Fragen
- Bilanz und Fazit

Unser Datensatz

- Pilotprojekt des im März 2021 eingereichten ANR-DFG-Projekts *TADS* (*Translation of Audio Description Scripts*)

AKTUELLE FOLGEN



- 24 Folgen der Serie *Neues aus Büttow* (Eberlein 2017), im Rahmen eines INALCO-Praktikums annotiert mit XML-TEI (P5) & TreeTagger

Audiodeskription (AD)

10:23:29

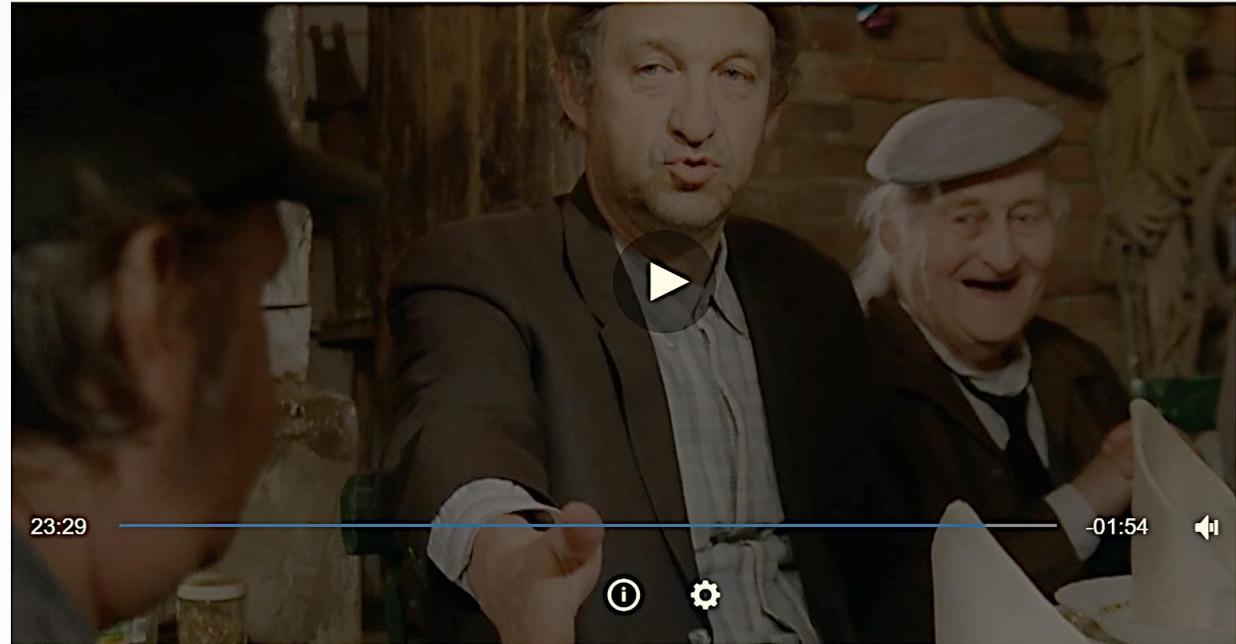
"Ja, herzlichen Glückwunsch."

ss Adsche gibt Brakelmann die Hand.

10:23:44

"Oh, eine Rede. Nee, dann lieber singen."

Genervt setzt Schönbiehl sich wieder.



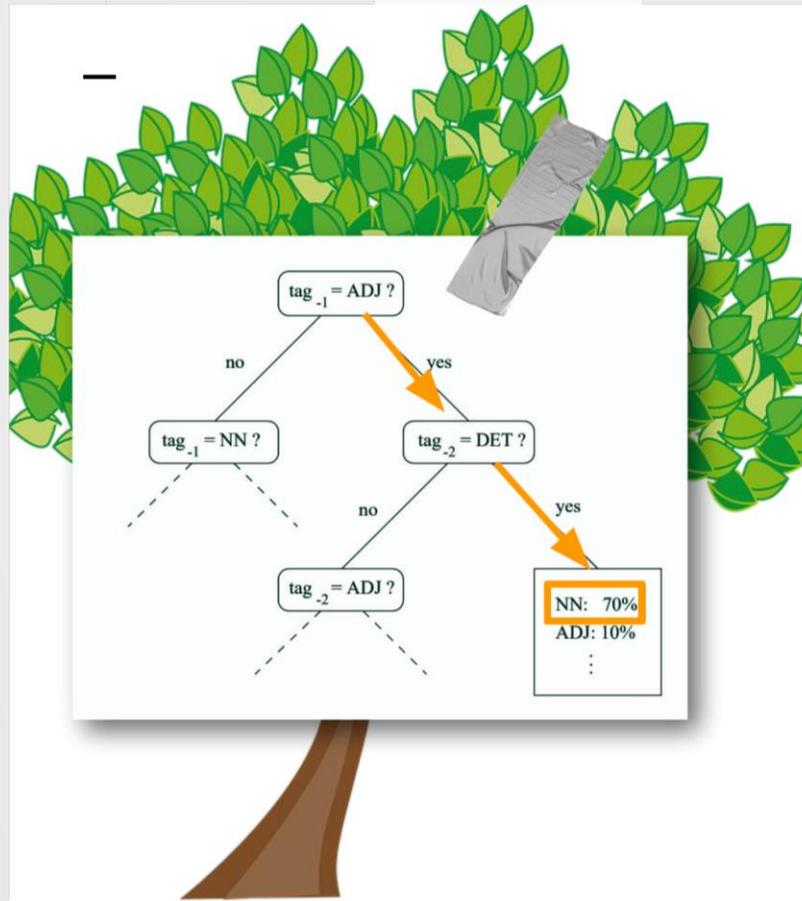
(Auszug des AD-Skripts der Folge 60 von *Neues aus Büttow*)

XML-Etiketten (AD-Sektionen)

AD-Sektion	Etikett XML-TEI	Beispiel
Aufzunehmender Text <i>speaker</i>	<sp>	Brakelmann fährt durchs Dorf.
Didaskalien (Handlungsanweisungen bezüglich der Aufnahmen)	<stage>	(Rest übersprechen)
Sprechtempo	<stage type="delivery">	s, ss, n [schnell, sehr schnell, normal]
Dialogbruchstück (vor Beginn des aufzunehmenden Texts)	<prompt>	"Goethe gehört mir!"
Schrift im Film	<caption>	Schild am Straßenrand. "Goethe Eier 5 Euro 10"

XML-Etiketten der Sektionen *speaker*, *stage*, *prompt*, *caption*.

TreeTagger



Decision Trees

Gesucht: $P(\text{NN}|\text{DET}, \text{ADJ})$

= 0,7

- Sobald zweimal ja gesagt wird, wird der Wahrscheinlichkeitsvektor ausgegeben
- Beispiel STTS: Maximal 106 Tests für 54 Tags
- Entscheidungsbäume werden durch einen Algorithmus berechnet, der die Tests so wählt, dass möglichst wenig Schritte gemacht werden müssen

Quelle:
[Deborah Watty](#)

Adverbien und Adverbiale im Stuttgart-Tübingen-Tagset (STTS 1.0.)

3.7 Adverbien

3.7.1 ADV: “echte” Adverbien

Als Adverbien werden nur reine, nicht von Adjektiven abgeleitete, nicht flektierbare Modifizierer von Verben, Adjektiven, Adverbien und ganzen Sätzen verstanden. Wortformen, die auch als attributive Adjektive auftreten und adverbial verwendet werden, die aber semantisch nichts (mehr) mit dem Adjektiv verbindet, und die meistens auch nicht prädikativ verwendet werden können, werden zu den Adverbien gezählt (z.B. *nämlich*).

Klassifikation von ADV

POS

POS =	Beschreibung	Beispiele
ADV	lokale Adverbien temporale Adverbien modale Adverbien kausale Adverbien Abtönungspartikel Präp. + “einander” Ordinalzahlen Multiplikativzahlen abgekürzte Formen	<i>dort, da, fort</i> <i>heute, dann, oft</i> <i>gerne, sehr</i> <i>darum, sonst</i> <i>ja, aber, denn, doch, zwar</i> <i>miteinander, nebeneinander</i> <i>erstens, zweitens, drittens</i> <i>einmal, zweimal, dreimal</i> <i>u.a., z.B.</i>
Aber:		
ADJD	adverbial gebrauchtes Adjektiv adverbial gebrauchtes Partizip Perfekt	<i>er fährt schnell/ADJD,</i> <i>ein schlecht/ADJD <i>gespieltes Stück</i></i> <i>er fährt gekonnt/ADJD</i>

Adverbien und Adverbiale im STTS 1.0.

3.2.2 ADJD: prädikativ oder adverbial gebrauchte Adjektive

Klassifikation von ADJD		POS
POS =	Beschreibung	Beispiele
ADJD	“echte” Adjektive (Positiv) (Komparativ, Superlativ) ursprüngliche Nomina adverbial gebrauchtes Partizip Präsens adverbial gebrauchtes Partizip Perfekt Ordinalzahlen	<i>[sie ist] groß, [es ist] lila</i> <i>[er läuft] schneller/am schnellsten</i> <i>[es ist] recht</i> <i>[er kam] lachend [herein]</i> <i>gekonnt [gespielt]</i> <i>[schneller als] geplant</i> <i>[sie sind zu] zweit</i>
Aber:		
ADV	<u>nur</u> adverbial gebrauchte Form	<i>er kommt nämlich/ADV morgen</i>
CARD	Kardinalzahlen	<i>sie waren zwei/CARD</i>
VVPP	nicht flektiertes Partizip Perfekt	<i>er wird gesucht/VVPP,</i> <i>es ist geplant/VVPP</i>
PTKVZ	adjektivische abgetrennte Verbzusätze	<i>[er hält] geheim/PTKVZ</i>

Adverbien und Adverbiale im STTS 1.0.

3.2.3 ADJD oder VVPP?

- Partizipien in adverbialer Stellung: ADJD.

Beispiele:

- *er spielt **gekonnt**/ADJD*
- *er kommt **geflogen**/ADJD*
- *die Mittel wurden **gezielt**/ADJD **eingesetzt**/VVPP*

- Attributiv oder modifizierend verwendete Partizipien werden als ADJD getaggt, ebenso Partizipien nach *wie* und *als*.

Beispiele:

- *er macht es **wie geplant**/ADJD*
- *sie lügt **wie gedruckt**/ADJD*
- *die **geplante**/ADJA Sache*
- *das **gewollt**/ADJD **verlorene**/ADJA Spiel*

- Lexikalisierte Partizipien. Problemfälle sind Passivpartizipien (Vorgangspassiv: mit *werden*, Zustandspassiv: mit *sein*), die je nach Kontext auch eine adjektivische Lesart zulassen (z.B. *verrückt*: *Patiens* = [+BELEBT] ⇒ ADJD).

Beispiele:

- *der Tisch wird **verrückt**/VVPP*
- **aber:** *der alte Mann wird **verrückt**/ADJD*

Resultat einer TreeTagger-Analyse

```
2112 Oh ITJ
2113 , $,
2114 eine ART
2115 Rede NN
2116 . $.
2117 Nee NE
2118 , $,
2119 dann ADV
2120 lieber ADJD
2121 singen VVINF
2122 . $.
2123 " $(
2124 Genervt VVPP
2125 setzt VVFIN
2126 Schönbiehl NN
2127 sich PRF
2128 wieder ADV
2129 . $.
```

Auszug aus der Folge 60 des Korpus *Buettenwarder*

Falsche Etiketten:

lieber ADJD > ADV

Genervt VVPP > ADJD

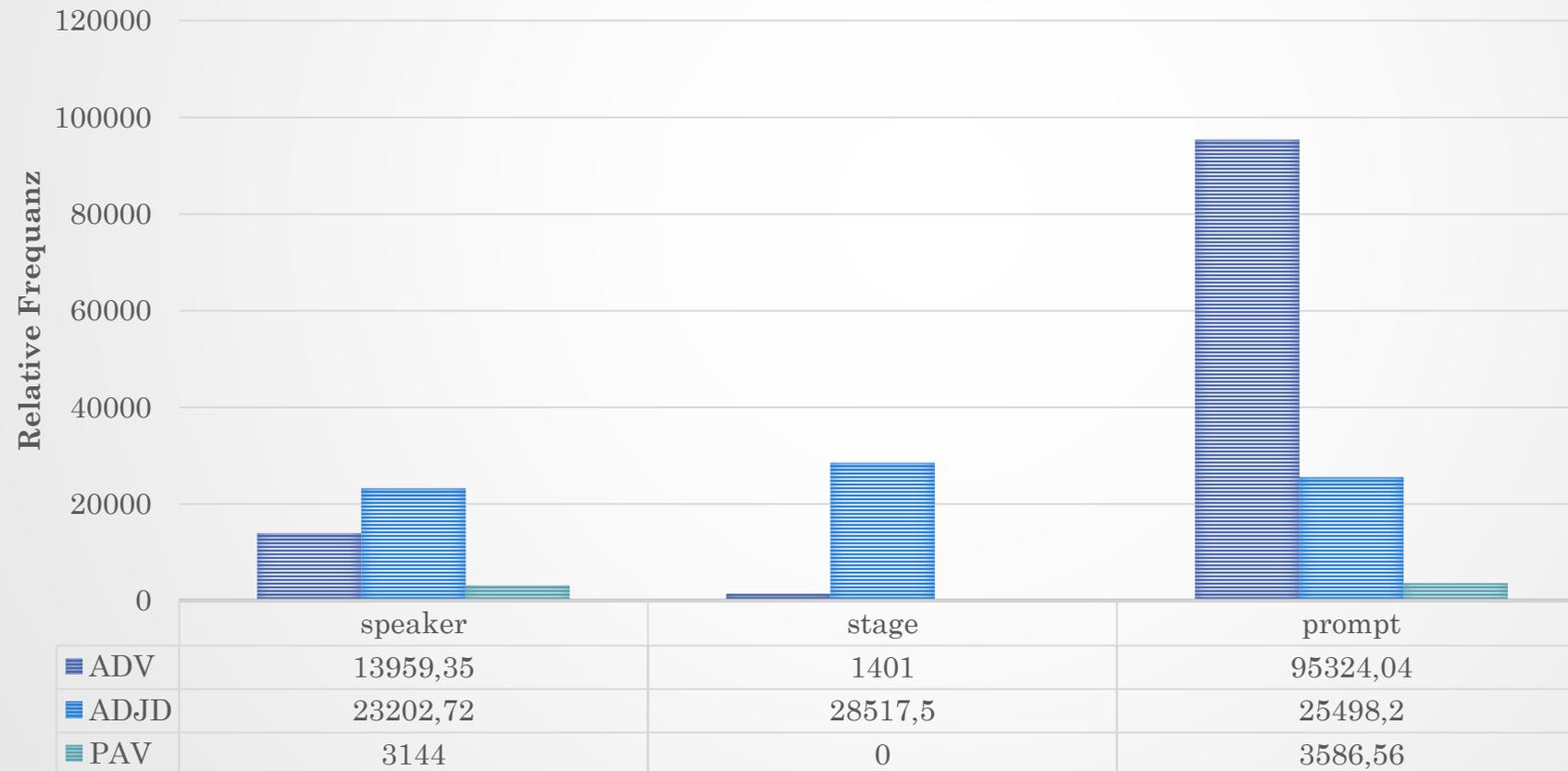
TreeTagger (Helmut Schmid, 1995)

TreeTagger-Etiketten für Adverbien und Adverbiale in AD-Skripten

Etikett	Definition	Beispiel
ADV	Adverb	schon, bald
ADJD	Adverbiales oder prädikatives Adjektiv	[er fährt] schnell
PAV	Pronominaladverb	dafür, dabei
ADVAD	AD-spezifisches Adverb	s, ss, n

Vergleich der Sektionen

VERTEILUNG ADV, ADJD, PAV



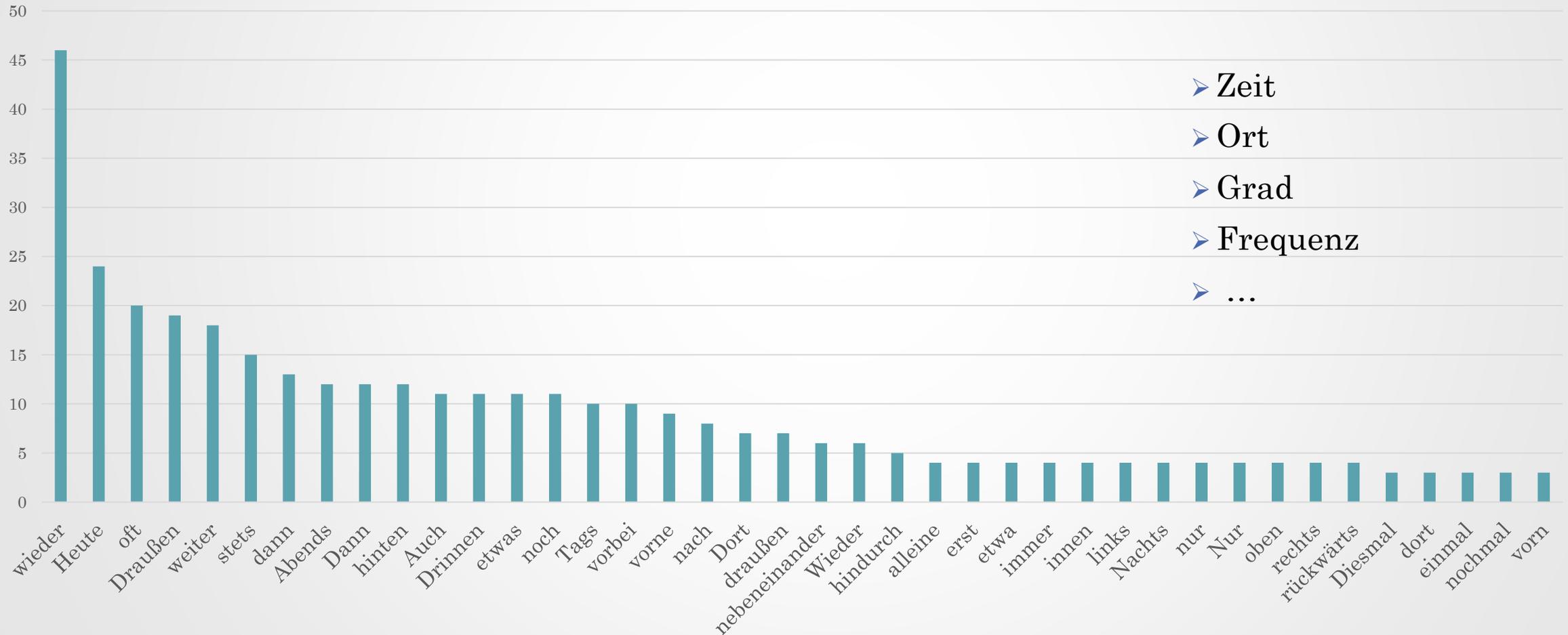
ADV in der Sektion *stage*

11: 50 (bisschen	Entengeräusch)	Eine Ente watschelt mit einem guten Dutzend
22: 51 (bisschen	Fegen)	Onkel Krischan und Adsche sitzen im zugeschneiten
22: 52 (bisschen	Trommelwirbel)	Shorty nimmt in der ersten Reihe Platz
15: 54 (bisschen	Motor) s	Ein dreirädriger Kleintransporter kommt. Adsche
22: 57 (bisschen	Duell-Musik)	Adsche hebt langsam mit links die Tomate
(Hupe 2	mal	frei)	Mit zusammengepressten Lippen setzt er sich wieder

- **bisschen:** <bisschen Motor>
 - **x mal:** <Hupe 2 mal frei>, <einmal TELEFONRINGEN>
 - **etwas:** <"Haar" etwas übersprechen>, <etwas Kopfhöreremusik>
- Grad
- Frequenz

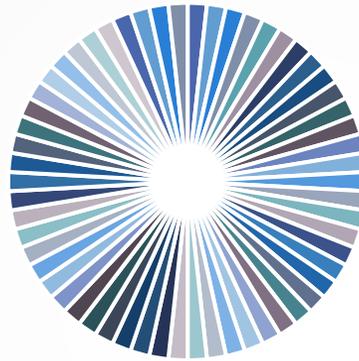
ADV in der Sektion *speaker*

Absolute Frequenz (T = 31799)



ADV in der Sektion *prompt*

Höchste Frequenzen (3810-7173 WpM):
ja, mal, so, auch, hier, doch, noch



- Zeit
- Ort
- Grad
- Frequenz
- ...

Hapaxformen:

■ Abend	■ Aber	■ alleine	■ bald	■ bestimmt	■ bisher
■ DA	■ DANN	■ Denn	■ donnerstags	■ dort	■ einfach
■ Endlich	■ etwas	■ fast	■ gegeneinander	■ Gerne	■ gestern
■ halt	■ Heut	■ hinaus	■ immerhin	■ je	■ jeher
■ jemals	■ längst	■ Mehr	■ miteinander	■ Morgen	■ nach
■ Nachmittags	■ Natürlich	■ Nebenan	■ neulich	■ Niemals	■ NOCH
■ nochmal	■ öfter	■ Rein	■ schließlich	■ selbst	■ Sofort
■ sonntags	■ Sonst	■ soweit	■ sowieso	■ stattdessen	■ überall
■ Überall	■ umsonst	■ unlängst	■ vorhin	■ woanders	■ Zuerst

Fehlerquote TreeTagger (STTS 1.0.)

ambiguity	tokens	in %	correct	in %	lexical errors	in %	disambig. errors	in %
0	1342	15.10	1128	84.05	214	15.95	0	0.00
1	5401	60.77	5330	98.69	71	1.31	0	0.00
2	993	11.17	929	93.55	3	0.30	61	6.14
3	795	8.95	757	95.22	0	0.00	38	4.78
4	260	2.93	240	92.31	0	0.00	20	7.69
5	96	1.08	83	86.46	0	0.00	13	13.54
total	8887	100.00	8467	95.27	288	3.24	132	1.49

Trainingsdaten:
Presstexte

(Watty 2017)

Fehlertypen

- Lexical errors: Das korrekte Tag ist nicht im Lexikon, und der Tagger rät ein inkorrektes Tag.
- Disambiguation errors: Das korrekte Tag ist im Lexikon, aber der Tagger wählt ein falsches Tag.

TreeTagger-Fehlerquellen (Folge 73)

479	Vor	APPR
480	Brakelmanns	NN
481	Füßen	NN
482	liegt	VVFIN
483	etwas	ADV
484	unter	APPR
485	einem	ART
486	Tuch	NN
487	.	\$.
488	#	XY
489	ss	VVFIN

Polykategorialität → Indefinitpronomen
versus adverbialer Gebrauch

Unbekannte Form <ss> → AD-spezifisches
Adverb wird hier als finites Verb analysiert

TreeTagger-Fehler ADJD

Form	Frequenz ADJD	Passendes Etikett
n	4	ADVAD
äh, Äh	2	ITJ
Dankeschön	4	PTKANT
dat	1	DET
(da)gegeng gekommen	1	VVPP
gegentreten	1	VVINF
GEHÖRT	1	VVPP
gewollt	1	VVPP
Griem	67	NE
hochverehrter	2	ADJA
is	7	VVFIN
los Los	23 3	PTKVZ
MACHEN	1	VVINF
[als] einziger	1	PIS?

Form	Frequenz ADJD	Passendes Etikett
mann	2	ITJ
Nee	1	PTKANT
nicht	19	PTKANT
Pe	15	NE
psst	1	ITJ
Radio-Raum	1	NN
rosa	2	ADJA
sach	5	VVFIN
ss	6	ADVAD
standst	1	VVFIN
Tomatenkartons	1	NN
wär, WÄR	4	VVFIN
zusammenfummeln	1	VVINF
hochverehrter...	2	ADJA

ADJD: adverbial oder prädikativ gebrauchte Adjektive

Analyse der ADJD-Fehler

Im Korpus *Buettenwarder* wird fälschlicherweise als ADJD gekennzeichnet, was...

- ungewöhnlich ist (mündliche Formen, Eigennamen, *speaker mode*)
- grafisch anderen Formen gleicht (ge...t, ge...n; etwas)
- ganz in Großbuchstaben geschrieben wird (MACHEN)
- syntaktisch isoliert erscheint ("Los!")

ADJD: adverbial oder prädikativ gebrauchte Adjektive

ADJD in der Sektion *speaker*

02: 38 Brakelmann	verdreht	genervt die Augen. Kuno stutzt
Kuno stutzt und lächelt	verlegen	. Schönbiehl kommt rein. 10
. Shorty sieht ihn	verärgert	an. Kuno schaut zu Brakelmann
Blick. Schönbiehl trinkt	betreten	von seinem Bier. Die drei
s Brakelmann sieht ihn	erbst	an. In seiner Küche.
Schnapsglas. Adsche guckt	betreten	vor sich hin. Er trägt
: 42 Adsche schaut	kurz	runter, dann wieder zu Brakelmann
mit der Hand.	Später	. Die Scheuentür fällt beim Öffnen
startet das Mofa.	Schlingend	fährt er vom Hof. Er
Kreide: 'Heute	geschlossen	'. Er dreht sich zu
: 59 Adsche grübelt	kurz	, dann sieht er hoch und
09: 05 s	Lächelnd	reicht Shorty ihm... " Ruf
06 Brakelmann sieht ihn	regungslos	an. 10: 11:
frei) s...	entzündet	Kuno eine Lunte. 10:
Scheune kommt Rauch.	Verstört	wanken Kuno und Brakelmann durch den
. Ihre Gesichter sind	rußgeschwärzt	. 10: 11: 35
: 09 Er schwingt	leicht	die Arme. 10: 12
steht auf und verschwindet	kurz	. Schönbiehl betritt den Schankraum.
" s Brakelmann dreht	langsam	den Kopf zu Adsche. 10
Musik kurz frei)	Lächelnd	richtet er sich auf, pustet
pustet durch und betrachtet	erneut	sein Spiegelbild in dem Silbertablett.
In Shortys Auto.	Lächelnd	richtet er seinen Hut. Der
macht Brakelmann den Raketenzyylinder	startklar	. Daran ist eine Holzlatte befestigt

ADJD > VVFIN

ADJD korrekt trotz syntaktischer Isolierung

Menschliche Analyse der Form "etwas"

(Studierende Master 1 TAL INALCO)

Beispiele	Passendes Etikett
Mürrisch steht Adsche vom Tresen auf und tritt etwas_ADV zurück.	
Brakelmann entdeckt etwas_ADJD am Boden.	PIS
hält einen seiner Gummistiefel umgedreht hoch und lässt etwas_ADJD in die Hand rollen.	PIS
"... ist etwas_ADJD Wunderbares!"	PIAT
Beim Trinken verschüttet sie etwas_ADV Köm.	
Er zupft etwas_ADJD von der Lippe...	PIS
Er entfernt sich etwas_ADV und bleibt stehen.	
("Haar" etwas_ADV übersprechen)	
Er beugt sich zu etwas_ADJD runter	PIS
Auf einem Teller präsentiert der Stämmige etwas_ADJD längliches Schwarzes.	PIAT
10: 23: 04 (etwas_ADV Kopfhörermusik)	
trottet der Bauer näher und starrt weiter auf etwas_ADJD hinter Kuno.	PIS
Vor einem der Ballen ist etwas_ADJD am Boden explodiert.	PIS
Heinzi schreibt am Tresen etwas_ADJD auf einen Bierdeckel.	PIS
Oder erkennt ihr auf dem Teller etwas_ADJD Anderes als Kekse?	PIAT
Vor Brakelmanns Füßen liegt etwas_ADJD unter einem Tuch.	PIS
Adsche zeigt etwas_ADJD zwischen Erde auf einem Tuch.	PIS
Adsche raunt 'Black Bronco etwas_ADV ins Ohr.	PIS
Anzahl der Analysefehler	13 von 18

PIS: substituierendes Indefinitpronomen; PIAT: attributierenden Indefinitpronomen (beide ohne Determiner)

Maschinelle Analyse der Form "etwas"

Contexte gauche	Pivot	Contexte droit
Mürrisch steht Adsche vom Tresen auf und tritt	etwas_PIS	zurück. Am Tisch sitzen Schönbiehl und Kuno vor Unterlagen. 10
" 10: 18: 45 Brakelmann entdeckt	etwas_ADV	am Boden. 10: 18: 49 " Was suchst du
hält einen seiner Gummistiefel umgedreht hoch und lässt	etwas_ADV	in die Hand rollen. Er betrachtet den Walnuss-großen Stein und wirft
10: 12: 44 "... ist	etwas_ADV	Wunderbares! " Jürgen nickt betrübt. 10: 12: 59
" Prost. " Beim Trinken verschüttet sie	etwas_ADV	Köm. Kuno fängt ihn mit der Hand auf und leckt ihn
44 "... Tatsächlich? " Er zupft	etwas_ADV	von der Lippe... 10: 26: 50 " Danke.
" Tarup Tarup ") Er entfernt sich	etwas_ADV	und bleibt stehen. 10: 19: 23 " Piep!
Sternen Haar. " (" Haar "	etwas_ADV	übersprechen) ss Adsche fährt. 10: 03: 34 "
Adsche suchend um. Er beugt sich zu	etwas_PIS	runter, greift es mit beiden Händen und zieht es über den
! " Auf einem Teller präsentiert der Stämmige	etwas_ADV	längliches Schwarzes. 10: 17: 36 " Wie sieht es
Brakelmann. 10: 23: 04 (etwas_ADV	Kopfhörermusik) ss Jürgens Finger nähert sich dem Kopfhörer des Stämmiger
trottet der Bauer näher und starrt weiter auf	etwas_PIS	hinter Kuno. Der dreht sich um und schaut in dieselbe Richtung
EXPLOSION) s Vor einem der Ballen ist	etwas_ADV	am Boden explodiert. Dichter grauer Rauch vernebelt kurz das Bild.
vor sich hin. Heinzi schreibt am Tresen	etwas_ADV	auf einen Bierdeckel. Shorty beugt sich verwundert zu ihm vor.
STUDIO: Oder erkennt ihr auf dem Teller	etwas_PIAT	Anderes als Kekse?] 10: 20: 50 " Ja
auf einem Radweg. Vor Brakelmanns Füßen liegt	etwas_ADV	unter einem Tuch. ss Im Dorfkrug zeigt Adsche Griem ein Modell
" – " Hier. " Adsche zeigt	etwas_ADV	zwischen Erde auf einem Tuch. 10: 11: 27 "
Anhänger. Adsche raunt 'Black Bronco '	etwas_ADV	ins Ohr und entfernt sich von Bruno und dem Pferd. (

Anzahl der Analysefehler: 12 von 18

PIS: substituierendes Indefinitpronomen; PIAT: attributierenden Indefinitpronomen (beide ohne Determiner)

Offene Fragen

Form	TreeTagger	Alternativer Vorschlag
Unter Bäumen hindurch _ADV blickt er zu Brakelmanns Hof.	ADV	?
Speisen werden zum Mund geführt – und nicht umgekehrt _ADJD .	ADJD	ADV?
"Ja und? Ich leb halt _ADV gut!"	ADV	PTKMA?
"Schönen Tach noch _ADV, ne. Man sieht sich."	ADV	PTKMA?
"Ja, ja. Das ist auch _ADV gut."	ADV	PTKMA?

Bilanz

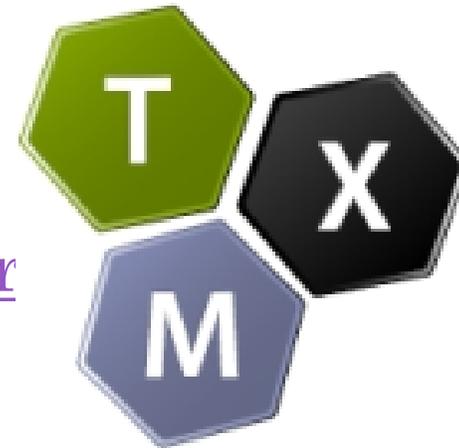
- Häufige Fehlerquellen der TreeTagger-Analysen unseres Datensatzes sind polykategoriale Formen und Formen, die dem TreeTagger unbekannt sind.
- Sowohl die automatische als auch die menschliche Analyse der Adverbien und Adverbiale, insbesondere die der Sektion *prompt*, die mündlichen Diskurs enthält, geben Anlass zur wissenschaftlichen Diskussion.
- Die Interpretation mancher der untersuchten Elemente ist besonders kontextabhängig.

Fazit

- Man sollte für weitere Analysen andere Annotierungswerkzeuge benutzen:
 - Mit Transkripten mündlicher Daten trainierte TreeTagger-Version STTS 2.0. (Westphal et al. 2017; siehe auch Beißwenger et alii 2015)
 - Helmut Schmid's [RNNTagger](#) : "a neural part-of-speech tagger and lemmatizer"
- Menschliche linguistische Analyse bleibt trotz maschineller Unterstützung unabdingbar.

Herzlichen Dank für Ihre Aufmerksamkeit!

- evalacroix@free.fr
- celine.largier-vie@sorbonne-nouvelle.fr



Bibliografische Angaben

- Beißwenger, M., Bartz, T., Storrer, A. & Westphal, S. 2015. Tagset und Richtlinie für das PoSTagging von Sprachdaten aus Genres internetbasierter Kommunikation. EmpiriST 2015.Shared Task des Empirikom-Netzwerks zur automatischen linguistischen Annotation deutschsprachiger internetbasierter Kommunikation. <https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/5065>
- Drosdowski, G. et al. (1995). Grammatik der deutschen Gegenwartssprache, Duden Band 4 (5. Auflage), Mannheim, Leipzig, Wien, Zürich: Dudenverlag
- Eberlein, N. (1997). Neues aus Büttenwarder [Fernsehserie]. https://www.ndr.de/fernsehen/sendungen/neues_aus_buettenwarder/folgen/index.html
- Geuder, W. (2019). Eine Art Wortart: Das Adverb im Deutschen. Zeitschrift für Sprachwissenschaft, 38(2), 191–241. <https://doi.org/10.1515/zfs-2019-2004>
- Heiden, S., Magué, J.-P., & Pincemin, B. (2010). TXM: Une plateforme logicielle open-source pour la textométrie - conception et développement. Proceedings of 10th International Conference Journées d'Analyse statistique des Données Textuelles, 2(3), 1021–1032. <https://halshs.archives-ouvertes.fr/halshs-00549779/document>
- Schaeffer-Lacroix, E., & Berland, K. (eingereicht 2021). Dealing with variation in audio descriptions scripts. *JAT (Journal of Audiovisual Translation)*, 10.
- Schiller, A., Teufel, S., & Stockert, C. (1995). Vorläufige Guidelines für das Tagging deutscher Textcorpora mit STTS. https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/stts_guide.pdf
- Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. Proceedings of International Conference on New Methods in Language Processing. International Conference on New Methods in Language Processing, Manchester, UK. <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf>
- Watty, Deborah. TreeTagger (Powerpointpräsentation vom 24.11.2017, Proseminar Tools für Computerlinguist/innen, Alexander Koller, Universität Saarbrücken). <https://coli-saar.github.io/tools17/slides/POS-Tagging.pdf>
- Westpfahl, S., Schmidt, T., Jonietz, J., & Borlinghaus, A. (2017). *STTS 2.0. Guidelines für die Annotation von POS -Tags für Transkripte gesprochener Sprache in Anlehnung an das Stuttgart Tübingen Tagset (STTS)*. <https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/6063>