

Zur Realisierung von pWortgrenzen in der gesprochenen Sprache

Pia Bergmann (Universität Freiburg)

„Empirical approaches to the phonological structure of words“

Marburg, 30. Oktober 2014

Das theoretische Konzept: pWort

- **Interface-Domäne zwischen Morphologie und Phonologie** (e.g. Booij 1985, Nespor & Vogel 1986 /2007, Hall 1999, Raffelsiefen 2000, Szczepaniak 2007, Wiese 2000)
- **Silbifizierung**
 - täg.+lich, schad.+los
 - kin.d+isch, far.b+ig, Gewerkschaf.t+er
 - Ur.+oma, Un.+art, Ver.+antwor.t+ung
 - Zahn.+arzt, Stand.+uhr
- **Phonologische Prozesse**
 - [Uŋgarn]_w
 - [un]_w[gern]_w *oder* [uŋ]_w[gern]_w
 - [Schrift]_w[tum]_w [glaub]_w[bar]_w
 - [tantst]_w von /tants/ + /st/ (2. Ps. Sg.)

Auslautverhärtung

**Glottalverschluss-
epenthese**

Assimilation

Degeminierung

Allerdings...

- Die meisten Analysen basieren auf Introspektion (cf. Hall & Kleinhenz 1999, Wiese 2000, Raffelsiefen 2000)
- Grenze und pWortstatus nicht notwendigerweise kategorisch (z.B. Szczepaniak 2007, Turk 2010)
- Einflussfaktoren auf Grenzstärke und pWortstatus über Sprechstil und Sprechgeschwindigkeit hinaus?
 - Frequenz-basierte Maße (z.B. relative frequency *softly/soft* vs. *swiftly/swift*, Hay 2003; cf. Bell et al. 2009, Bush 2001, Bybee 2001, Bybee & Scheibmann 1999)
 - Morphologische Informativität (z.B. in *+igheid*, Pluymaekers et al. 2010)
 - Semantische Transparenz / Ausbleichung? (Booij 1999, Giegerich 1985)
 - Lexikalischer Status? (z.B. Hall 1999, Wiese 2000)
- Relevanz anderer phonologischer oder phonetischer Eigenschaften, z.B. Dauer (z.B. Auer 2002, Pluymaekers et al. 2010, Turk 2010)

Die empirischen Blicke

Produktionsstudie I

- Degeminierung
- Glottalverschluss / Glottalisierung
- akustisch-phonetisch
- Partikelverben und Komposita

Produktionsstudie II

- Velarnasalassimilation
- artikulatorisch (EPG)
- Partikelverben und Komposita

Korpusstudie I: Erstglieder

- t-Tilgung
- Dauern
- Spontansprache
- akustisch-phonetisch
- Derivation – Komposition
ent- Grund-/Haupt- Fund-

Korpusstudie II: Letztglieder

- Dauern
- Spontansprache
- akustisch-phonetisch
- Derivation – Komposition
*-ig / -isch / -er -los -lohn
-lich / -ler*

Die empirischen Blicke

Produktionsstudie I

- Degeminierung
- Glottalverschluss / Glottalisierung
- akustisch-phonetisch
- Partikelverben und Komposita

Produktionsstudie II

- Velarnasalassimilation
- artikulatorisch (EPG)
- Partikelverben und Komposita

Korpusstudie I: Erstglieder

- t-Tilgung
- Dauern
- akustisch-phonetisch
- Derivation – Komposition
ent- Grund-/Haupt- Fund-

Korpusstudie II: Letztglieder

- Dauern
- akustisch-phonetisch
- Derivation – Komposition
-ig / -isch / -er -los -lohn
-lich / -ler

Empirie: Produktionsstudie „Degeminierung“

- n#n, l#l, s#s, t#t
- kontrollierte Wörter, eingebettet in Trägersätze
- 14 Sprecher aus Nordwestdeutschland (7w, 7m; Alter 19-35)

n#n				
	Höhere Frequenz		Niedrige Frequenz	
	VV	V	VV	V
Komp	Bahn#netz	Brenn#nessel	Bahn#nutzung	Spann#netz
PV	ein#nehmen ein#nisten	hin#nehmen	ein#nagen ein#nageln	hin#neigen

(Bahn)_ω(netz)_ω – (ein)_ω(nehmen)_ω

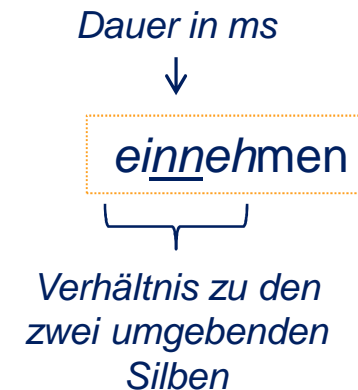
Empirie: Produktionsstudie „Degeminierung“

- eingebettet in Trägersätze
- IP-Position und Akzentuierung kontrolliert

akzentuiert	(Wie äußert sich deine Allergie gegen das Medikament?) Wenn ich es EINnehme, bekomme ich MAGENkrämpfe.
unakzentuiert	(Warum schluckst du ein Medikament, das du nicht gut verträgst?) Wenn ich es NICHT einnehme, geht es mir NOCH schlechter.

Empirie: Produktionsstudie „Degeminierung“

- Abhängige Variablen
 - absolute Dauer (log) der Lautsequenz
 - relative Dauer (log) der Lautsequenz
- Unabhängige Variablen
 - Tokenfrequenz (log; COSMAS II)
 - Wortart (Kompositum – Partikelverb)
 - Vokalquantität (V – VV)
 - Segment (n#n, l#l, s#s, t#t)
 - Akzentuierung (akzentuiert – unakzentuiert)
- Statistik: Lineares mixed-effects Regressionsmodell (cf. Baayen 2008)



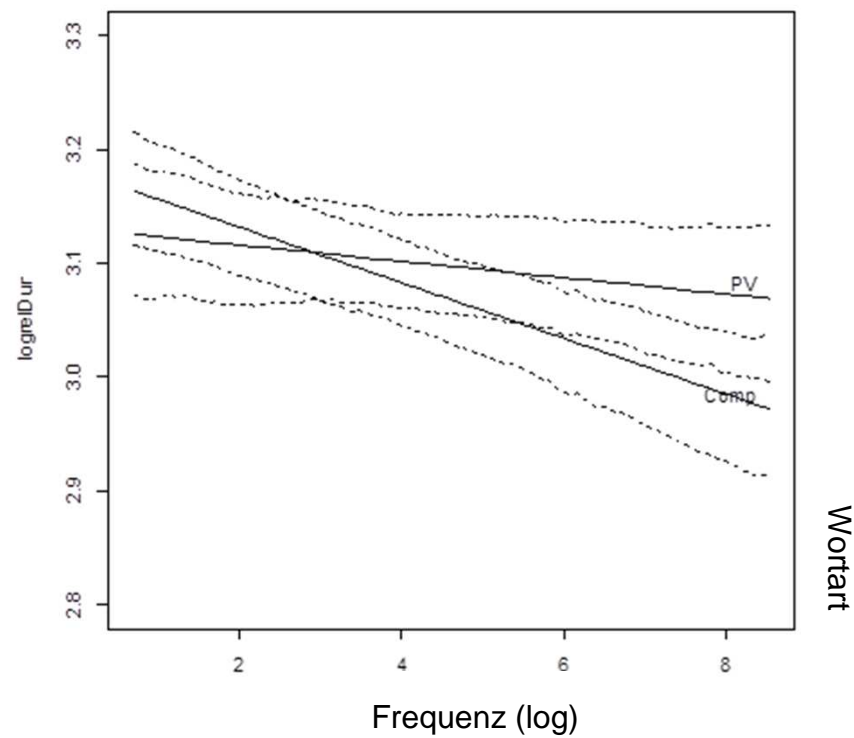
Produktionsstudie „Degeminierung“

- Ergebnisse

	Frequ	Lex	Frequ* Lex	akz	V	Segment	Frequ* Segment	
Komposita (n = 668) / R ² = 0.72	*	--	--	*	*	*	*	abs. Dauer
Komposita + PV (n = 1083) / R ² = 0.69	*	n.s.	*	*	*	*	*	
Komposita + PV reduziertes Set (n = 999) / R ² = 0.70	*	n.s.	n.s.	*	*	*	*	
Komposita + PV reduziertes Set (n = 997) / R ² = 0.70	*	(*) (t-Wert -1.98)	* (t-Wert 3.61)	*	*	*	*	rel. Dauer

Produktionsstudie „Degeminierung“

- Modellplot Wortart*Frequenz



Produktionsstudie „Degeminierung“

- Zusammenfassung: Dauern der Lautsequenz

- | | |
|-------------------|---------------------------------------|
| • Tokenfrequenz ✓ | LF > HF |
| • Akzentuierung ✓ | akz > unakz |
| • Wortart (✓) | PV weniger beeinflusst durch Frequenz |
| • Vokalquantität | V > VV |
| • Segment | Obstruenten > Sonoranten |

- graduelle Reduktion bei Geminaten existiert, auch über pWortgrenzen hinweg
- Variation ist systematisch und teilweise auf gebrauchts- und wortbezogene Faktoren zurückzuführen

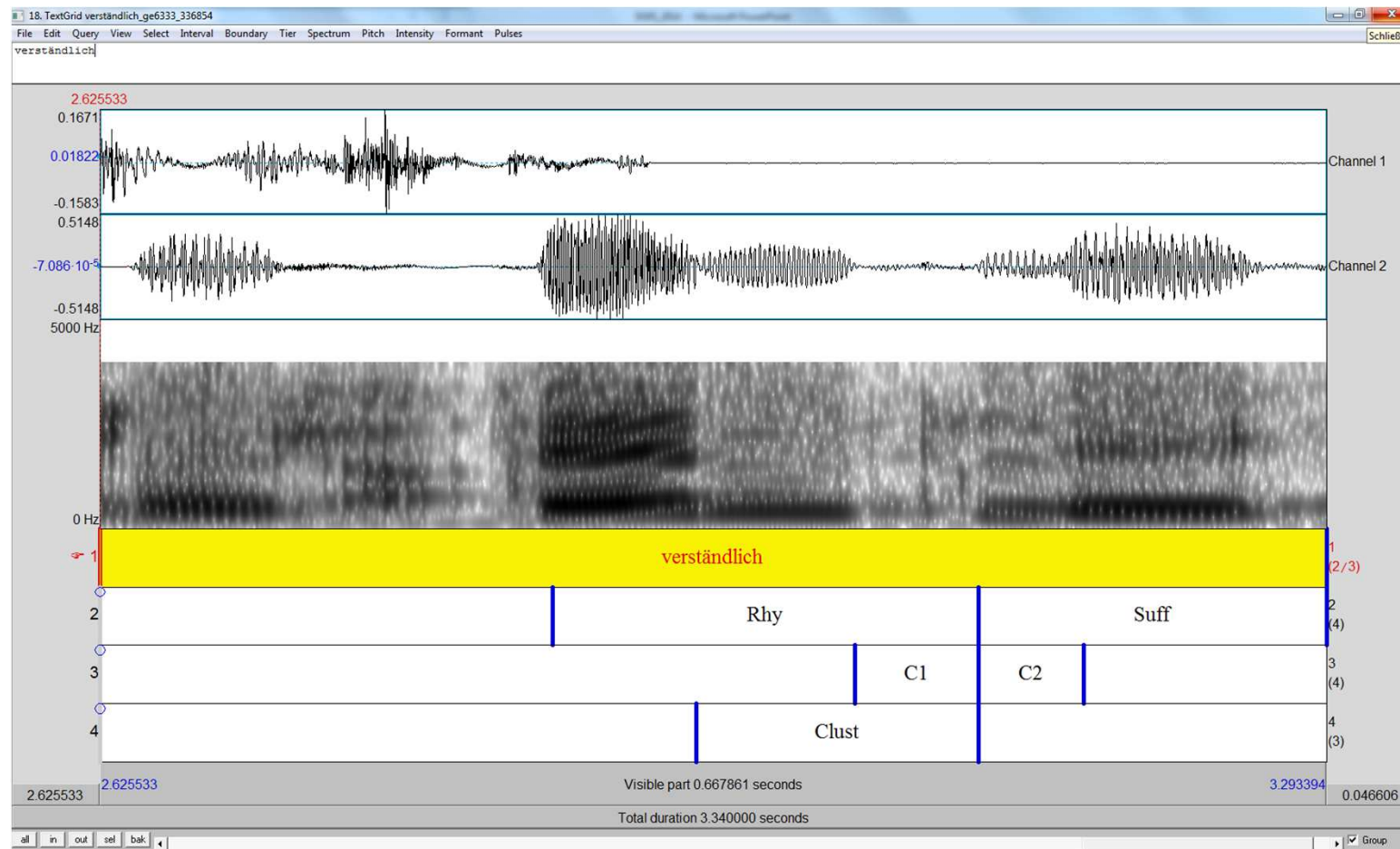
Korpusstudie „Letztglieder“

C-initiale vs. V-initiale Suffixe

- spontansprachliche Daten
- Datenüberblick

SUBSET	Beispielwörter		pWort	n = 610
Suffixe	h ä n d	l + e r	eins	103
	p o s t +	l e r		79
	s c h w i n d	l + i g		86
	e η	l + i s c h		120
	e n d +	l i c h	zwei	222
	C1	C2		
	Clust			
	Rhy			
Messungen				

Ein Beispiel



Hypothesen

Abhängige Variablen jeweils

- Dauern

- **Hypothese „pWort“**

- 2 pWörter > 1 pWort
- *+lich* > *l+ig, l+isch, l+er, +ler*

- **Hypothese „Frequenz“**

- absolute Lexemfrequenz
- LF > HF
- *Ständler* > *Händler*

- **Hypothese “Semantische Transparenz”**

- transparent > intransparent
- *stündlich* > *endlich*

Statistische Analyse

- lineares gemischtes Regressionsmodell (Baayen 2008)
- **Kovariate**
 - Reimstruktur und Artikulationsart C1
 - metrische Struktur im Wort
 - Satzakzent und IP-Position
 - Sprechgeschwindigkeit
 - random factor: Sprecher

Ergebnisse für die // -Dauer (C2)

(schwindlig)_ω vs (stünd)_ω(lich)_ω

Random effects:

Groups	Name	Variance	Std.Dev.
Sprecher	(Intercept)	2.6947e-05	0.0051911
	Residual	2.0086e-04	0.0141723

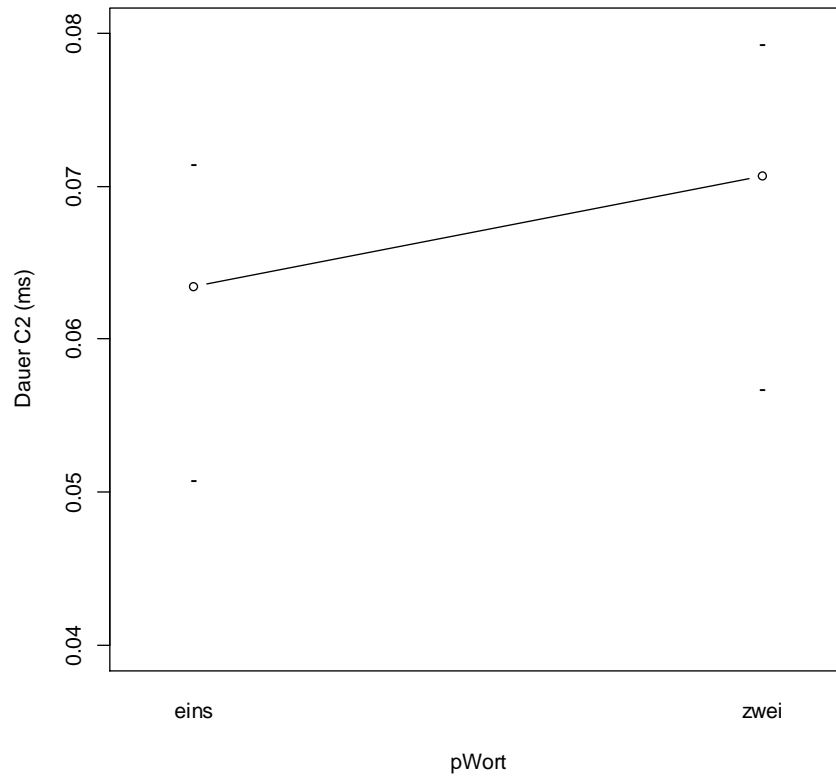
Number of obs: 545, groups: Sprecher, 194

Fixed effects:

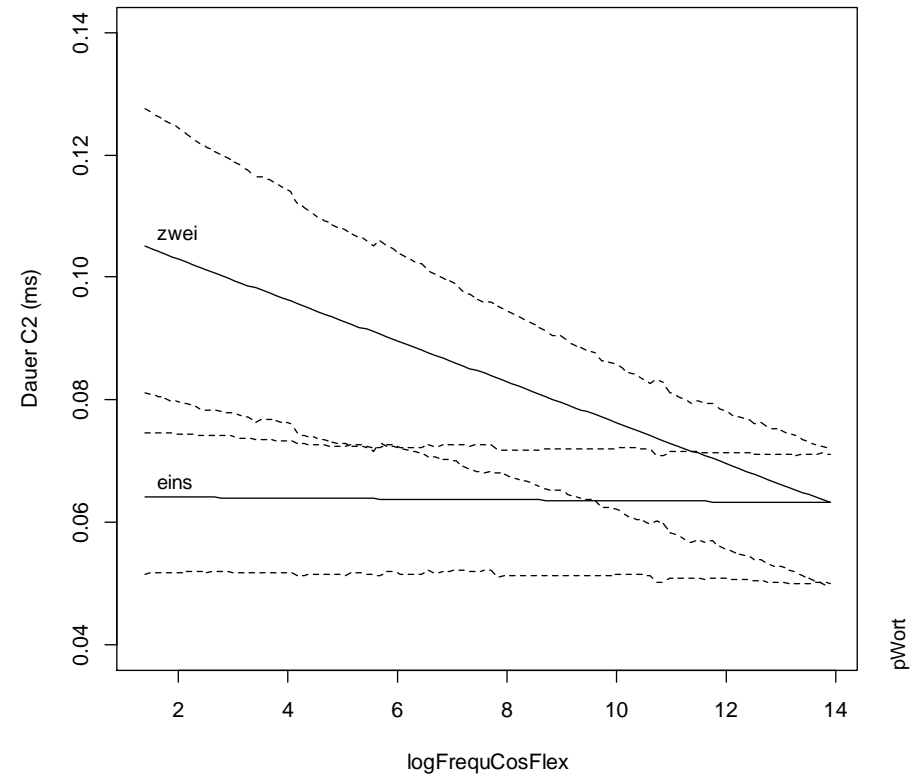
	Estimate	Std. Error	t value	
(Intercept)	8.861e-02	7.471e-03	11.859	
pWort zwei	4.535e-02	1.277e-02	3.551	*
Frequenz	-7.196e-05	3.374e-04	-0.213	
SemTrans nein	3.854e-03	2.494e-03	1.545	
pWort zwei:Frequenz	-3.264e-03	1.050e-03	-3.107	*
pWort zwei:SemTrans nein	-1.133e-02	3.412e-03	-3.320	*
ReimVC	-1.299e-02	5.238e-03	-2.481	*
ReimVCC	-1.813e-02	5.278e-03	-3.434	*
ReimVCCC	-2.574e-02	5.815e-03	-4.427	*
ReimVV	-4.199e-03	5.125e-03	-0.819	
ReimVVC	-1.910e-02	5.454e-03	-3.502	*
ReimVVCC	-1.179e-02	8.621e-03	-1.368	
ReimVVCCC	-1.224e-02	1.161e-02	-1.053	
Metrikww	-1.574e-02	7.185e-03	-2.191	*
IP_Posmedial	-9.175e-03	1.358e-03	-6.756	*
SilSec	-3.889e-03	6.381e-04	-6.095	*
SilSec:Metrikww	2.299e-03	1.134e-03	2.028	*

Ergebnisse: Einflussfaktoren pWort und Frequenz

1 pWort < 2 pWörter
(t-Wert = 3.551)



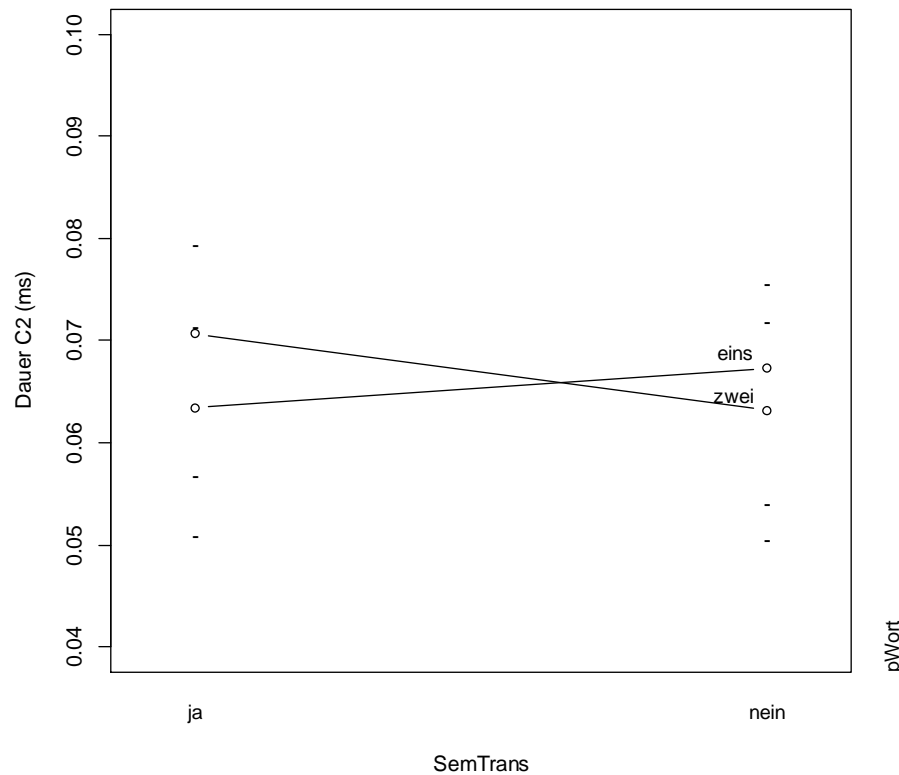
pWort * Frequenz
(t-Wert = -3.107)



Ergebnisse:

Faktoren pWort und semantische Transparenz

*pWort * semantische Transparenz*
(t-Wert = -3.32)



zwei: *stündlich* vs. *endlich*
eins: *Postler* vs. *Ständler*

		SemTrans		
		ja	nein	ges.
pWort	eins	313	48	361
	zwei	81	113	194
	ges.	394	161	

Korpusstudie „Letztglieder“ / C-initial vs. V-initial

- Zusammenfassung: Dauer des C2
 - pWort ✓ 2 Wörter > 1 Wort
 - Tokenfrequenz ✓ LF > HF
 - semantische Transparenz⁰ transparent > intransparent

- pWortstruktur relevant
 - höhere Dauern bei komplexen pWörtern
 - Frequenz beeinflusst komplexe pWörter, aber nicht einfache pWörter

Korpusstudie „Erstglieder“

- spontansprachliche Daten

		Beispielwörter
Präfixe (n = 245)	ent-	<i>entkleiden</i> <i>entsetzen</i>
„Präfixoide“ (n = 266)	Grund- Haupt-	<i>Grundwissen</i> <i>Grundsatz</i>
Erstglieder in Komposita (n = 194)	Grund- Fund- Bund- ...	<i>Grundbesitz</i> <i>Bundhose</i> <i>Rundfunk</i>

„Affixoid“-Kriterien

- reihenbildend
- semantisch ausgebleicht
- gleichlautendes freies Lexem
(vgl. Munske 2002, Stein 2008)

Korpusstudie „Erstglieder“

- spontansprachliche Daten

		Beispielwörter	Semantik	pWort
Präfixe (n = 245)	ent-	<i>entkleiden</i> <i>entsetzen</i>	+ trans - trans	umstritten
„Präfixoide“ (n = 356)	Grund- Haupt-	<i>Grundwissen</i> <i>Grundsatz</i>	+ trans, + distanz - trans, + distanz	zwei
Erstglieder in Komposita (n = 148)	Grund- Fund- Bund- ...	<i>Grundbesitz</i> <i>Bundhose</i> <i>Rundfunk</i>	+ trans, - distanz + trans, - distanz - trans, + distanz	zwei

Korpusstudie „Erstglieder“

- Unabhängige Variablen
 - Tokenfrequenz
 - Typefrequenz des Erstglieds
 - Semantik
 - morphologische Kategorie
- Abhängige Variablen
 - t-Tilgung
 - Dauer des finalen Clusters (/nt/, /pt/)
 - abs. und rel. Dauer des grenzübergreifenden Clusters (/Ct#C/)

Hypothesen

Abhängige Variablen jeweils

- Dauern (finales Cluster, grenzübergreifendes Cluster)
- t-Tilgung

- **Hypothese „Frequenz“**

(1) Tokenfrequenz

- $LF > HF$

(2) Typefrequenz

- $LF > HF$

Hypothesen

Abhängige Variablen jeweils

- Dauern (finales Cluster, grenzübergreifendes Cluster)
- t-Tilgung

- **Hypothese “Semantik”**

(1) Distanz / Ausbleichung

- keine Distanz > Distanz

(2) Transparenz

- transparent > intransparent

- **Hypothese „Morphologische Kategorie“**

- Erstglieder in Komposita > „Präfixoide“ > Präfixe

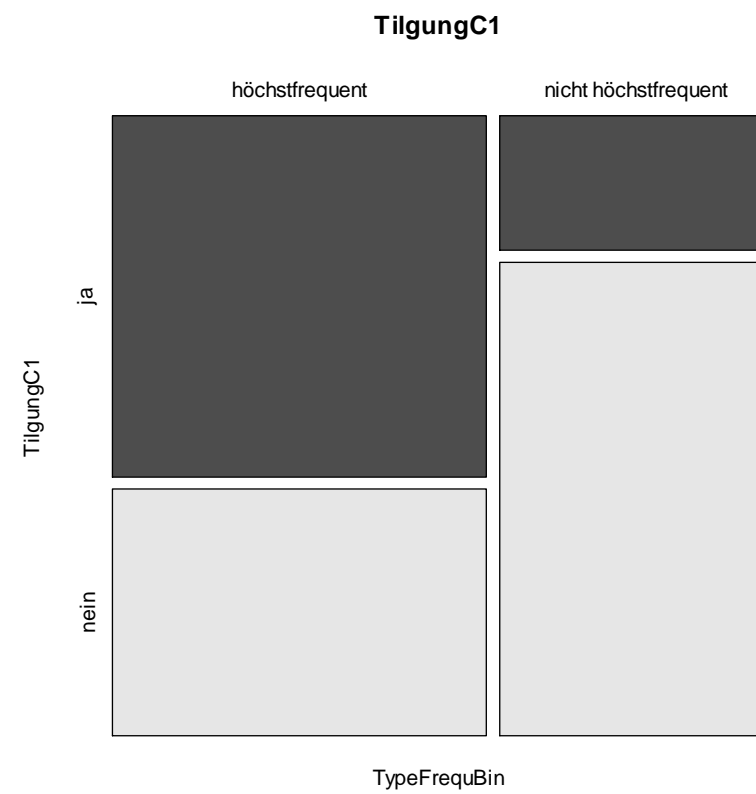
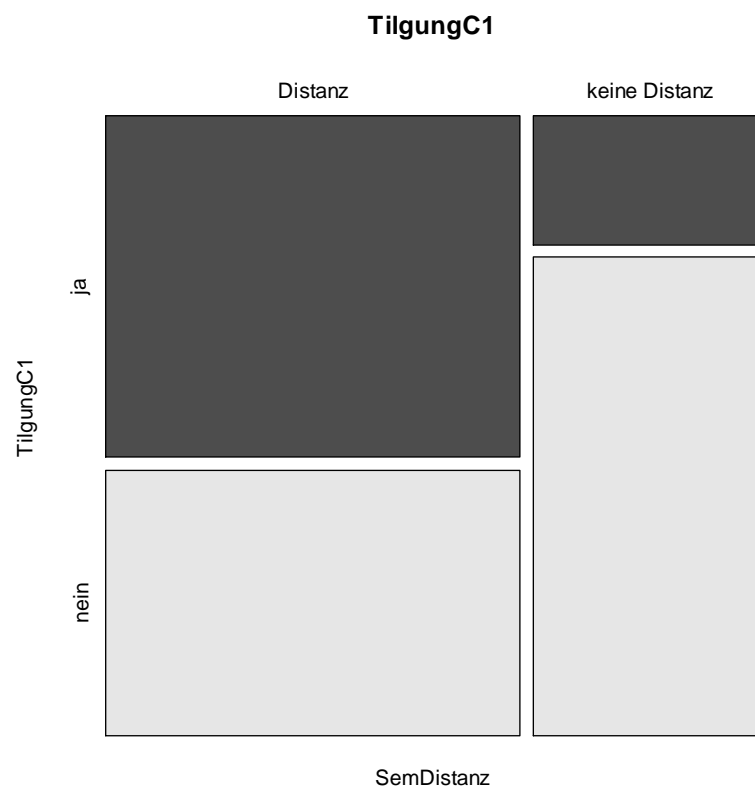
Kovariate

- Cluster (-nt, -pt)
- Folgesegment (Konsonant vs. Vokal)
- Phonotaktik (legal vs. illegal, z.B. *Haupttraum*, *Grundbesitz*)
- Segmentkontext erstes Segment im Cluster (Assim: *Grundbesitz*, Geminate: *Hauptttbahnhof*, Affrikate: *Hauptftfahrzeit*)

- Wortakzent (Präfixe vs. andere; aber auch einige bei *grundsätzlich* und *hauptsächlich*)
- IP-Position (final vs. nicht-final)
- Silbenanzahl/Wort
- Sprechgeschwindigkeit (Silben/Sekunde)
- Region
- Sprecher

/t/-Tilgung

- Rohdatenplots



Ergebnisse /t/-Tilgung (nur für /nt/)

- **Hypothese „Frequenz“**
 - (1) Tokenfrequenz (✓)
 - LF > HF
 - (2) Typefrequenz ✓
 - LF > HF
- **Hypothese “Semantik”**
 - (1) Distanz / Ausbleichung (✓)
 - keine Distanz > Distanz
 - (2) Transparenz ⁰
 - transparent > intransparent
- **Hypothese „Morphologische Kategorie“**
 - Erstglieder in Komposita > „Präfixoide“ > Präfixe ⁰

Ergebnisse /t/-Tilgung (nur für /nt/)

- **Hypothese „Frequenz“**

(1) Tokenfrequenz (✓)

- LF > HF

(2) Typefrequenz

- LF > HF

- **Hypothese „Sprechgeschwindigkeit“**

(1) Distanz / Ausk

- keine Distanz >

(2) Transparenz 0

- transparent > i

- **Hypothese „M**

- Erstglieder in k

Random effects:

Groups	Name	Variance	Std.Dev.
Sprecher	(Intercept)	1.3757e+00	1.1729e+00
Region	(Intercept)	1.0594e-10	1.0293e-05

Number of obs: 237, groups: Sprecher, 70; Region, 8

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.79796	1.33324	2.849	0.00439 **
Frequenz	-0.18367	0.07979	-2.302	0.02134 *
SemDistanz: keine	0.70532	0.38731	1.821	0.06860 .
C2: Vokal	2.20887	1.31054	1.686	0.09190 .
Sprechgeschwindigkeit	-0.18103	0.16839	-1.075	0.28235

/t/-Tilgung

Random effects:

Groups	Name	Variance	Std.Dev.
Sprecher	(Intercept)	6.9516e-01	8.3376e-01
Region	(Intercept)	6.3263e-13	7.9538e-07

Number of obs: 426, groups: Sprecher, 91; Region, 8

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.2127	0.8285	0.257	0.7974

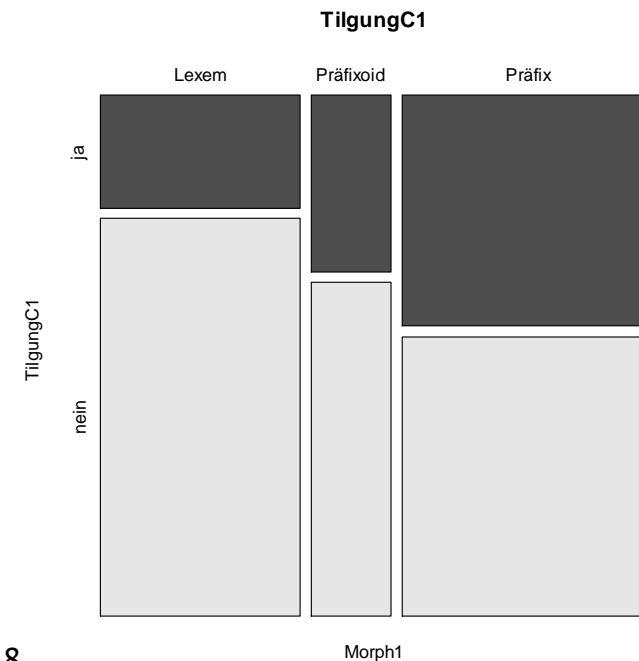
Morph1: Lexem

Morph1: Präfixoid

C2: Vokal

Sil/Wort

Sprechgeschwindigkeit



- „Präfixoide“ gruppieren sich mit den Präfixen
- die sie konstituierenden Eigenschaften (hohe Typefrequenz und semantische „Ausbleichung“) sind auch für sich genommen signifikant

Zusammenfassung

- systematische Clusterreduktion in Abhängigkeit von
 - Tokenfrequenz
 - Typefrequenz
 - semantischer Ausbleichung
- Wortakzent
- segmenteller Kontext (Clustertyp, Folgesegment)

Fazit

- systematische Clusterreduktion in Abhängigkeit von
 - Tokenfrequenz
 - Typefrequenz
 - semantischer Ausbleichung
 - Wortakzent
 - segmenteller Kon
- die Komplexität der pWortgrenze unterliegt systematischer Reduktion
 - sie ist nicht unbeeinflusst durch Folgeelemente im angrenzenden pWort
 - die umstrittenen Präfixe sind tatsächlich stärkerer Reduktion unterworfen
 - der gebrauchts- (und wort-)bezogene Faktor „Frequenz“ spielt eine signifikante Rolle

Vielen Dank für die Aufmerksamkeit!