Frequency effects and prosodic boundary strength

Tina Bögel

University of Konstanz

w.i.p.

in cooperation with Alice Turk, University of Edinburgh

Saarbrücken, 22.1.2020

Overview

- The *Smooth Signal Redundancy Hypothesis*
- Experiment on frequency effects with respect to
  - the placement of syntactic boundaries
  - the strength of the resulting prosodic boundaries

  → durational measurements of the boundary-related intervals
- Frequency effects and grammar architecture
- Conclusion

The smooth signal redundancy hypothesis (SSRH)

- Often no reliable cues to indicate word boundaries in spoken language
- **Assumption**: prosodic boundary structure is planned to achieve SSR (Aylett and Turk 2004, Turk 2010)

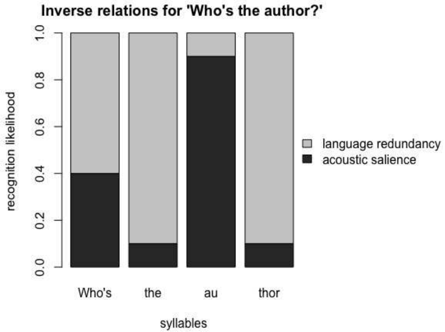
→ make the recognition of each word in an utterance equally likely

→ prosodic boundary strength assumed to **inversely** relate to language redundancy, i.e., non-acoustic information:

  - likelihood syntactic structure
  - lexical word frequency
  - word bigram frequency
  - ...

- More predictable elements require "less explicit signal information" than less predictable elements for successful recognition (Lindblom 1990)

Inverse relation



(Aylett 2000, Aylett and Turk 2004)

→ Inverse, complementary relationship between language redundancy and acoustic redundancy

→ Recognition likelihood spread evenly throughout an utterance

⇒ achieve maximal understanding with minimal effort

(Some) previous work

- More likely to pronounce (phrase-medial) syllables with low language redundancy more clearly (Aylett (2000), Aylett & Turk (2004, 2006))
- Jurafsky et al. (2001): highly frequent function words and function words with a high probability given context are more likely to be acoustically reduced
- Bell et al. (2009) showed an effect on word duration given the following material
- Pluymaekers et al. (2005) showed an effect of bigram frequency on stem and suffix duration; they also showed an effect of repetition
- Gahl & Garnsey (2004) showed that syntactic predictability can also affect segment and pause durations with transitive verbs
- Watson et al. (2006) showed that the likelihood of intonational boundary insertion was greater when the presence of a word's dependent was optional (less predictable) than when it was judged to be obligatory (more predictable)

Previous work

In conclusion, previous work showed that increased

- lexical frequency (e.g., Jurafsky et al. 2001)
- bigram frequency (e.g., Aylett 2000, Aylett and Turk 2004, 2006, Pluymaekers et al. 2005, Bell et al. 2009)
- syntactic predictability (e.g., Gahl and Garnsey 2004, Watson et al. 2006)

led to a reduction of word/segment duration, and influenced the placement of syntactic boundaries.

Clearly demarcating word boundaries → more salience

Hypothesis

- Inverse relationsip between language redundancy and acoustic salience should hold for larger prosodic boundaries

→ Stronger prosodic boundaries are expected to occur where language redundancy is low, e.g., within infrequent stretches of speech

- SSRH would thus predict a (gradient) correlation between boundary strength and language redundancy (e.g. greater final lengthening, initial lengthening, initial strengthening, F0 reset, etc., given low language redundancy)

→ Has not been tested experimentally!

Work presented here

- Investigates the relationship between language redundancy and prosodic boundary strength
- through the effect of:
  - syntactic frequency
  - word frequency
  - word bigram frequency

→ on the placement of intonational phrase boundaries

→ on durational measurements of boundary strength

Challenge:

Need to vary language redundancy, while using controlled material

- with similar syntactic phrasing
- with similar segments across boundaries (effects might be subtle)

Experimental design: syntactic ambiguities

When the cake was dropped flat plants stuck to its underside

- Syntax A: *the cake was* **dropped** .... **flat plants** *stuck to its underside* (= modifying construction, [V [A N]])
- Syntax B: *the cake was* **dropped flat** .... **plants** *stuck to its underside* (= resultative construction, [[V A] N])

Estimation of syntactic frequency

- The following corpora were used to estimate syntactic frequencies

	Brown corpus	**ICE-GB**
Released	1964	1998
Tagging	Part of Speech (POS)	Syntactic (Treebank)
Tokens	˜ 1 Million	˜ 1 Million
English	BE	AE
Texts	Across all genres	Edited English prose
Citation	(Francis and Kučera 1964)	(*ICE-GB corpus* 1998)

Table: Information on the ICE-GB and the Brown corpus

Experimental design: syntactic ambiguities II

Frequency determination:

	Verb-Adj		**Adj-Noun**
	main	copula	
ICE-GB corpus	1771	8781	21183
	10552		
In %	˜ **5%**	˜ 28%	˜ **67%**
	˜ 33%		
Brown corpus	1657	4562	47830
	10552		
In %	˜ **3%**	˜ 8,5%	˜ **88,5%**
	˜ 11,5%		

Table: Frequency of syntactic combinations in the ICE-GB and the Brown corpus

Conclusion:
Syntax A (=modifying) is far more likely than Syntax B (=resultative)

Experimental design: placement of phrase boundaries

- Difference in syntax comes with difference in the placement of an intonational phrase boundary

V **%** A N

or

V A **%** N

- Expect **V%AN** to occur more often (if speakers are given a choice)

→ corresponding syntactic structure is more frequent

Experimental design: lexical frequencies I

In order to determine:

1. effects of frequency **on syntactic choice**, the relevant syntactic sequence had to have four combinations:

Verb	Adj.	Noun	Shortcut
$V_{frequent}$	Adj.	$N_{frequent}$	ff
$V_{frequent}$	Adj.	$N_{infrequent}$	fi
$V_{infrequent}$	Adj.	$N_{frequent}$	if
$V_{infrequent}$	Adj.	$N_{infrequent}$	ii

2. effects of frequency **on boundary strength**, the four combinations above had to be comparable:
   - in the rhyme/coda of the verb
   - in the onset of the noun
   - in the onset and the rhyme/coda of the adjective

   → known to show the largest durational effects of boundary strength

   But: had to allow for reliable measurements at the same time

Experimental design: lexical frequencies II

Estimation of lexical frequencies via WebCelex:

	Verbs	**Nouns**
frequent	> 2000	> 3000
infrequent	< 200	< 100

Table: Raw number thresholds for lexical (in)frequencies

→ Matching of verbs/nouns with respect to the form

ff: *When the cake was d***ropped** *flat* **p***lants stuck to its underside*

fi: *When the cake was d***ropped** *flat* **p***lanks stuck to its underside*

if: *When the grass was c***ropped** *flat* **p***lants were able to grow again*

ii: *When the grass was c***ropped** *flat* **p***lanks were laid across the lawn*

Experimental design: lexical frequencies III

Examples with four combinations:

freq Verb	**infreq Verb**	**freq Nouns**	**infreq Nouns**
dropped	cropped	plank	plant
buy	dye	paper	paisley
call	wall	door	dorm
made	shade	picture	pitcher
make	rake	field	fief
stayed	bayed	sister	sissy
play	slay	fish	fiend
shake	snake	boxes	bobbers
turned	churned	balls	baulks
wear	pare	farmers	farthings
works	lurks	markets	marshals
walk	stalk	people	peafowls

Bigram frequencies

Determined **bigram frequencies** of Verb-Adj (V-A) and Adj-Noun (A-N) combination and their **ratio**: V-A/A-N

Problem: No corpus large enough to determine frequencies of infrequent combinations.

→ Google

→ 'Noisy', therefore just approximations

→ Great variance

⇒ Divided data into abstract categories:

	low	**med** (buffer)	**high**
	$< 40\%$	40% - $< 60\%$	$>= 60\%$
V_A	< 13900	< 314000	$>= 314000$
A_N	< 3180	< 108000	$>=108000$

Table: Abstract representation of raw bigram frequencies

Data gathering

- Data presentation:

Num	Block
1.	**without comma, repetition 1 (58)**
2.	with comma (112)
3.	without comma, repetition 2 (58)

Table: Presentation of sentences: 228 in total

→ without commas (syntactic boundary placed according to choice)

→ several repetitions; only discuss first repetition here (58 sentences/speaker)

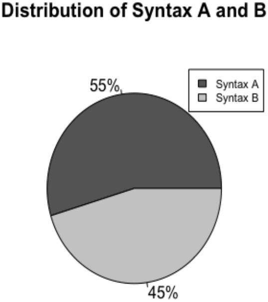
- Subjects: 23 participants
  (students at the University of Edinburgh, Ø=23,4 years, 14 females)
- Recordings: sound-treated studio at the University of Edinburgh with a high quality microphone

Frequency and syntactic choice: results I

Annotation of syntactic choice:
1 annotator (100%), 1 annotator (40%) – 100% agreement

Here: 23 speakers, repetition 1 → total of 1314 instances

Syntax A → V%AN
Syntax B → VA%N



→ almost equally distributed – surprising given the results from the corpora ....

Frequency and syntactic choice: results II

For the choice of syntax, the following factors were relevant:

- **Syntax A** (frequent syntax, V%AN) more likely with
  - highly frequent nouns ($p < 0.05$)
  - high A-N bigram frequency ($p < 0.001$)

- **Syntax B** (infrequent syntax, VA%N) more likely with
  - highly frequent verbs ($p < 0.001$)
  - high V-A bigram frequency ($p < 0.001$)
  - higher V-A in comparison to A-N bigram frequency ($p < 0.001$)

Durational measurements: preparation

Strict selection:

- Only speakers that generally had a high consistency across repetitions (1 sentence - 1 choice - in both repetitions)

→ 10 speakers

- Only quadruplets that had the **same syntactic choice across both repetitions**

→ can measure frequency impact on duration – and later compare it to repetition 2

- Today: Discuss only repetition 1

Annotated sentences	
Syntax A	*Syntax B*
124	54

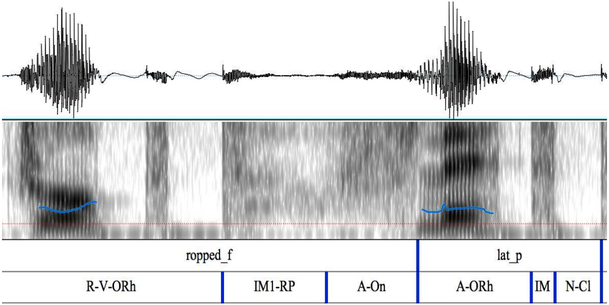
Durational measurements: annotation

- Raw material, e.g.

V-A	A-N	example
`ropped_f`	`lat_t`	*dropped flat plants*
`k_f`	`ree_p`	*walk free people*

→ Problematic, a lot of segmental variation

- Abstract annotation scheme, three intervals per sequence (six in total)



Durational measurements: annotation

Verb end		**Adjective start**		**Adjective end**		**Noun start**	
`V-Rh`	rhyme	`A-On`	onset	`A-Rh`	rhyme	`N-On`	onset
`V-Co`	coda	`A-Cl`	closure	`A-Co/Co1/Co2`	coda/coda part 1/2	`N-Cl`	closure
`V-ORh`	with part of onset			`A-ORh`	with part of onset		
`R-V-...`	with onset release			`A-Nu`	nucleus, not coda		
				`R-A-...`	with onset release		

Intermediate (`IM1` and `IM2`)		**Comment:**
`...-R`	release	*Might include aspiration!*
`...-P`	pause	*Missing pause (P) is only indicated if syntax requires it*
`...-RP`	release and pause	*Both -P/-RP are only indicated if there is no closure following*
		If no R/P is present and not expected, then leave out IM. Else use brackets ()

Supra-markers		**Comment:**
`?`	insecurity	*Insecurity in annotation, mostly at preceding or following border*
`x_x`	connection	*Connection across word boundaries - e.g., V-Rh_IM1_A-On*
`( )`	missing element	*For elements that should be there, but are not (mostly R and P)*
`NA`		*If a separation at word boundary in DurationSep (only!) is not possible*
`rel`	release	*Only on DurationSep level. Connected to other parts with +*
`pause`	pause	*Same as release*
`glot`	glottalization	*Same as release*
`(breath)`	non-expected release	*Same as release*

→ Allows for grouping of similar patterns to get more reliable measurements!

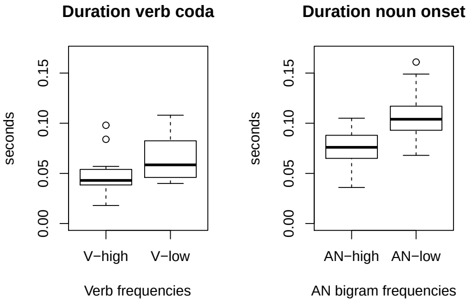
BUT: If there was no clear boundary, intervals were connected via an underscore (_)

→ particular item then not part of analysis - further reduction of data

Frequency and duration: some (significant) results I

Syntax A (frequent, V%AN):

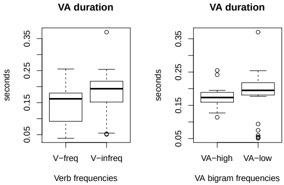
- When lexical frequency V is low: **increased verb coda interval duration** ($p < 0.05$)
- When bigram frequency AN is low: **increased noun onset interval duration** ($p < 0.05$)



Frequency and duration: some (significant) results II

Syntax B (infrequent, VA%N):

- When lexical frequency V is low or bigram frequency VA is low:
  increase overall VA duration
  ($p < 0.05$ and $p < 0.01$)



→ Same effect is found with the verb coda (but not with the adjective onset)
($p < 0.05$ and $p < 0.01$)

- When VA bigram frequency higher than AN frequency:
  - **decrease of verb coda interval duration**
    ($p < 0.001$)
  - **increase of noun onset interval duration**
    ($p < 0.01$)

Conclusion (duration)

All of these results are consistent with the SSRH:

- inverse relationship between language redundancy (lexical frequencies, bigram frequencies, and their interaction) and durational measurements of the prosodic boundary-related intervals

→ frequency effects are found to influence all levels of language

Question:

How can we encode this in a formal implementation of grammar?

... also with respect to possible uses in, e.g., speech synthesis applications

'performance' knocking at the door of *'competence'*

Prosodic structure and its interfaces

Two aspects to this question:

1. The likelihood of the syntactic structure
   - → Formally, this can be achieved via so-called OT-preference-marks in the syntactic encoding
   - → In the case of syntactically ambiguous structures, the **syntactic choice can be signalled via prosodic boundary placement**
   - ⇒ The prosodic boundary either occurs before (V%AN) or after (VA%N) the adjective

2. Inverse relationship of language redundancy and duration
   - → **Low frequency correlates with longer duration and vice versa**

We can formally analyze these processes in Lexical-Functional Grammar (LFG)

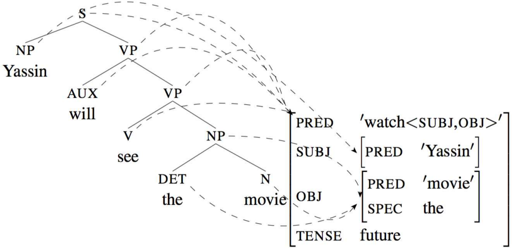
LFG - projection architecture

- LFG has a modular *projection architecture*.

→ The different levels of representation are related to each other via mathematically defined projections.

→ E.g., each piece of the c(onstituent)-structure contributes information to the f(unctional)-structure.

(2) Yassin will watch the movie.



c-structure **f-structure**

LFG's Projections

Over the years, more projections than the original core c-structure and f-structure have been argued for:

- s(emantic)-structure
- a(rgument)-structure: place for thematic roles and information about predicate composition (complex predicates)
- i(nformation)-structure: place for information structural components
- p(rosodic)-structure: information on intonation and on prosodic constituency

LFG's Projections

The architecture of LFG allows for complex interactions across projections.

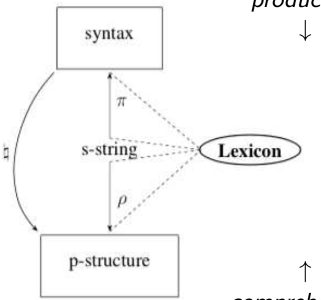
- Initial LFG proposals for the p-structure were "syntactocentric"
- Newer proposals have moved to seeing prosody as a separate level of representation that interacts with morphosyntax, but is not completely derived from it
- The analysis presented here is based on the syntax-prosody interface for LFG developed in Bögel (2015).

The Prosody-Syntax interface (Bögel 2015)

Two perspectives:

(Roughly following models as proposed by, a.o., Levelt (1999) and Jackendoff (2002)

- *Production*: from meaning to form (syntax → prosody)
- *Comprehension*: from form to meaning (prosody → syntax)



production

↓

↑

comprehension

♮: The *Transfer of structure* → Information on (larger) syntactic and prosodic phrasing, and on intonation is exchanged

ρ: The *Transfer of vocabulary* → Associates morphosyntactic and phonological information on lexical elements and projects them to their respective structures

P-structure – the p-diagram (during production!)

- Linear representation in the p-diagram

→ structured syllablewise

⇒ Each syllable is part of a vector associating the syllable with relevant values:
→ *lexical stress, segments, prosodic phrasing, ...*

- Includes language-specific phonological processes (postlexical phonology, prosodic restructuring)

PHRASING	$(\sigma)_\omega)_\iota$	$(_\iota(\sigma)_\omega$	$(\sigma)_\omega$	...
...	...	...	...	...
LEX_STRESS	prim	prim	prim	...
SEGMENTS	/drɒpt/	/flæt/	/plɑŋks/	...
V. INDEX	$\mathbf{S}_1$	$\mathbf{S}_2$	$\mathbf{S}_3$	$\mathbf{S}_4$

- **Input** to the p-diagram comes from syntax/c-structure (*Transfer of structure*) and the lexicon (*Transfer of vocabulary*)

The Transfer of Vocabulary

- Associates morphosyntactic and phonological information on lexical elements

→ Via the multidimensional lexicon, which projects them to their respective structures

s(yntactic)-form				**p(honological)-form**	
dropped	V	(↑ PRED)	= 'drop⟨SUBJ⟩'	P-FORM	[drɒpt]
		(↑ TENSE)	= past	SEGMENTS	/d r ɒ p t/
		...		METR. FRAME	$('\sigma)_{\omega}$
planks	N	(↑ PRED)	= 'plank'	P-FORM	[plɑŋks]
		(↑ PERS)	= 3	SEGMENTS	/p l ɑ ŋ k s/
		(↑ NUM)	= Pl	METR. FRAME	$('\sigma)_{\omega}$
		...			

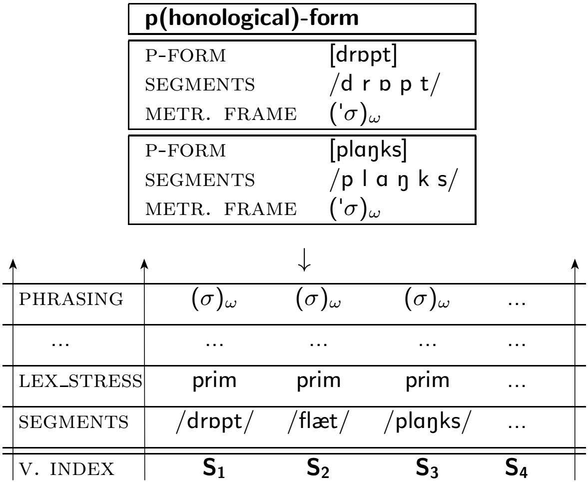
- Each lexical dimension can only be accessed by the related module

→ Modular: strict separation of module-related information

→ Translation function: Once a dimension is triggered, the related dimensions can be accessed as well.

⇒ Associated **p-form is selected and made available to p-structure**.

The Transfer of Vocabulary II



p(honological)-form

P-FORM	[drɒpt]
SEGMENTS	/d r ɒ p t/
METR. FRAME	(ˈσ)$_{\omega}$

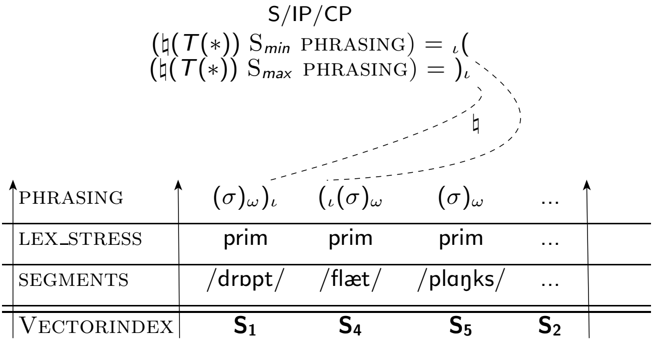
P-FORM	[plɑŋks]
SEGMENTS	/p l ɑ ŋ k s/
METR. FRAME	(ˈσ)$_{\omega}$

PHRASING	$(\sigma)_{\omega}$	$(\sigma)_{\omega}$	$(\sigma)_{\omega}$	...
...	...	...	...	...
LEX_STRESS	prim	prim	prim	...
SEGMENTS	/drɒpt/	/flæt/	/plɑŋks/	...
V. INDEX	$\mathbf{S}_1$	$\mathbf{S}_2$	$\mathbf{S}_3$	$\mathbf{S}_4$

- Also needed: Information on larger prosodic constituents

→ Via the *transfer of structure*

The Transfer of Structure ... from syntax to prosody



S/IP/CP

$$(\natural(T(*))\ \mathrm{S}_{min}\ \textsc{phrasing}) = {}_{\iota}($$
$$(\natural(T(*))\ \mathrm{S}_{max}\ \textsc{phrasing}) =)_{\iota}$$

PHRASING	$(\sigma)_{\omega})_{\iota}$	$(_{\iota}(\sigma)_{\omega}$	$(\sigma)_{\omega}$	...
LEX_STRESS	prim	prim	prim	...
SEGMENTS	/drɒpt/	/flæt/	/plɑŋks/	...
VECTORINDEX	**S_1**	**S_4**	**S_5**	**S_2**

- where S_{min} refers to the *first* syllable within the scope of a node
- where S_{max} refers to the *last* syllable within the scope of a node

→ Roughly following Selkirk (2011)'s *Match theory*

- But problem still unresolved: **Where should frequency effects be encoded?**

Frequency effects as part of the lexical entry

- Further dimension: *meta* information

→ Encodes the individual lexical frequency

→ Encodes bigram frequencies

s-form		**p-form**		**meta**	
dropped	V	P-FORM	[drɒpt]	LEX_FREQ	high
		SEGMENTS	/d r ɒ p t/	BI_FREQ	
		METR. FRAME	$(\text{ˈ}\sigma)_{\omega}$	drop=flat	high
				…	…
planks	N	P-FORM	[plaŋks]	LEX_FREQ	low
		SEGMENTS	/p l ɑ ŋ k s/	BI_FREQ	
		METR. FRAME	$(\text{ˈ}\sigma)_{\omega}$	flat=planks	low
				…	…

Frequency effects in the p-diagram

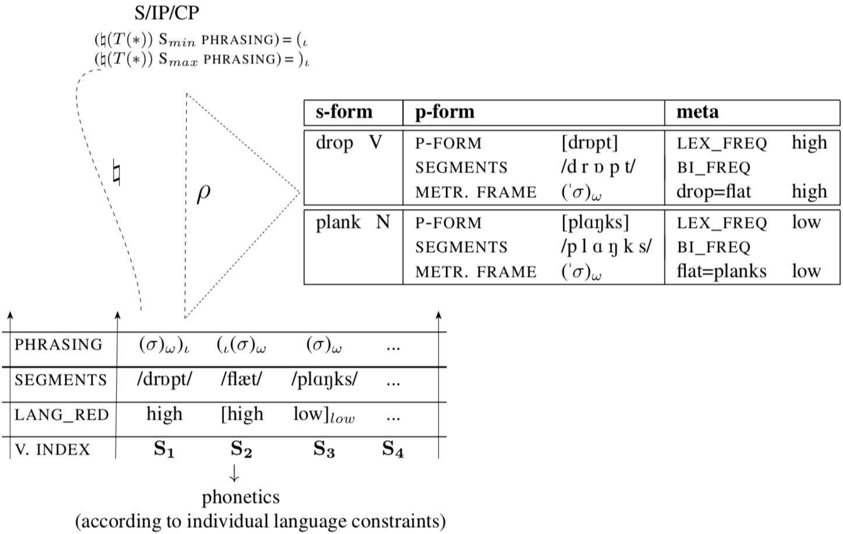
- Frequency information becomes part of the underlying representation, e.g. as LANGUAGE REDUNDANCY:

PHRASING	$(\sigma)_\omega)_\iota$	$(_\iota(\sigma)_\omega$	$(\sigma)_\omega$	...
SEGMENTS	/drɒpt/	/flæt/	/plɑŋks/	...
LANG_RED	(high	[high)$_{high}$	low]$_{low}$	...
VECTORINDEX	$\mathbf{S_3}$	$\mathbf{S_4}$	$\mathbf{S_5}$	...

- ... passed on to the phonology-phonetics interface ...
- where this information can be transformed into the associated acoustic cues

  → For syntax A:
  - When lexical frequency V is low: increased verb coda interval duration
  - When bigram frequency AN is low: increased noun onset interval duration

Overall framework



S/IP/CP

$(\natural(T(*))\ S_{min}\ \text{PHRASING}) = (_{\iota}$

$(\natural(T(*))\ S_{max}\ \text{PHRASING}) =)_{\iota}$

s-form	**p-form**		**meta**	
drop V	P-FORM	[drɒpt]	LEX_FREQ	high
	SEGMENTS	/d r ɒ p t/	BI_FREQ	
	METR. FRAME	$(^{\prime}\sigma)_{\omega}$	drop=flat	high
plank N	P-FORM	[plɑŋks]	LEX_FREQ	low
	SEGMENTS	/p l ɑ ŋ k s/	BI_FREQ	
	METR. FRAME	$(^{\prime}\sigma)_{\omega}$	flat=planks	low

PHRASING	$(\sigma)_{\omega})_{\iota}$	$(_{\iota}(\sigma)_{\omega}$	$(\sigma)_{\omega}$	...
SEGMENTS	/drɒpt/	/flæt/	/plɑŋks/	...
LANG_RED	high	[high	low]$_{low}$	...
V. INDEX	$\mathbf{S_1}$	$\mathbf{S_2}$	$\mathbf{S_3}$	$\mathbf{S_4}$

↓

phonetics

(according to individual language constraints)

Conclusion

- Language redundancy affects the strength of prosodic boundaries

→ Allows for a smooth signal: recognition of each element is equally likely

- Word and bigram frequencies are part of the lexical entry

→ This information can be encoded as part of the underlying p-structure

- Transformation into concrete acoustic cues at the interface between phonology and phonetics

Outlook:

- Compare repetitions
- Investigate F0
- Zoom in on bigram frequencies across boundaries
- ...

Thank you!

... questions, comments...?

References I

Aylett, Matthew. 2000. *Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and care of articulation in spontaneous speech*. Ph. D.thesis, University of Edinburgh.

Aylett, Matthew and Turk, Alice. 2004. The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Language and Speech* 47(1), 31–56.

Aylett, Matthew and Turk, Alice. 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America* 119(5), 3048–3058.

Bell, Alan, Brenier, Jason M., Gregory, Michelle, Girand, Cynthia and Jurafsky, Dan. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1), 92–111.

Bögel, Tina. 2015. *The Syntax–Prosody Interface in Lexical Functional Grammar*. Ph. D.thesis, University of Konstanz.

Francis, W. Nelson and Kučera, Henry. 1964. A Standard Corpus of Present-Day Edited American English, for use with Digital Computers. Technical Report, Department of Linguistics, Brown University, Providence, Rhode Island, US.

Gahl, Susanne and Garnsey, Susan. 2004. Knowledgef of grammar, knowledge of usage. Syntactic probabilities affect pronunciation variation. *Language* 80, 748–775.

ICE-GB corpus. 1998. ˙Online resource: http://ice-corpora.net/ice/icegb.htm .

Jurafsky, Dan, Bell, Alan, Gregory, Michelle and Raymond, William. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee and P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, pages 229–254, Amsterdam: John Benjamins.

Pluymaekers, Mark, Ernestus, Mirjam and Baayen, R. Harald. 2005. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118, 2561–2569.

Turk, Alice. 2010. Does prosodic constituency signal relative predictability? A Smooth Signal Redundancy hypothesis. *Laboratory Phonology* pages 227–262.

Watson, Duane G., Breen, Mara and Gibson, Edward. 2006. The role of syntactic obligatoriness in the production of intonational boundaries. *Journal of Experimental Psychology: Learning, Memory and Cognition* 32, 1045–1056.