Frequency effects and prosodic boundary strength

Tina Bögel

University of Konstanz

w.i.p.

in cooperation with Alice Turk, University of Edinburgh

Frankurt, 5.12.2018

Overview

- The *Smooth Signal Redundancy Hypothesis*
- Experiment on frequency effects and syntactic/prosodic boundaries
  - Experimental Design
  - Frequencies and syntactic choice
  - Frequencies and boundary strength
  - Conclusion
- Frequency effects and grammar architecture

The smooth signal redundancy hypothesis (SSRH)

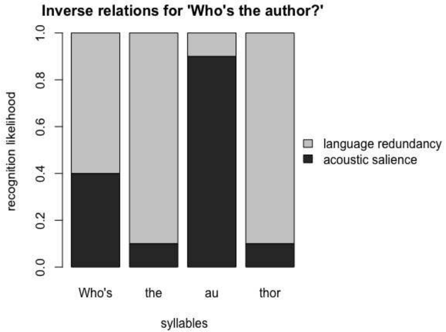
- Often no reliable cues to indicate prosodic boundaries in spoken language
- **Assumption**: prosodic boundary structure is planned to achieve SSR

→ 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)

Hypothesis

- Inverse relationsip between language redundancy and acoustic salience should hold for larger prosodic boundaries
- SSRH predicts greater final lengthening, initial lengthening, initial strengthening, F0 reset, etc., given low language redundancy

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

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

→ 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 I

When the cake was dropped flat plants stuck to its underside

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

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

V)ι(A N

or

A N)ι(N

Experimental design: syntactic ambiguities II

Frequency determination:

	Verb-Adj main	**Verb-Adj** copula	**Adj-Noun**
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 more likely than Syntax B (=resultative)

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

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 gras was* c**ropped** *flat* **p***lants were able to grow again*

ii: *When the gras 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**
	< 33%	33% - < 66%	>= 66%
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)

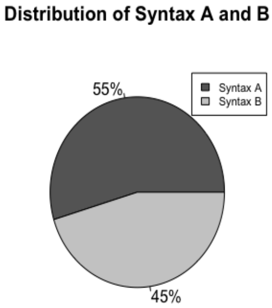
Table: Presentation of sentences: 228 in total

- 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



... surprising given the results from the corpora ....

Frequency and syntactic choice: results II

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

- More likely with **Syntax A** (frequent, [V [A N])
  - highly frequent nouns ($p < 0.05$)
  - high A-N bigram frequency ($p < 0.001$)
- More likely with **Syntax B** (infrequent, [[V A] N])
  - 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 (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

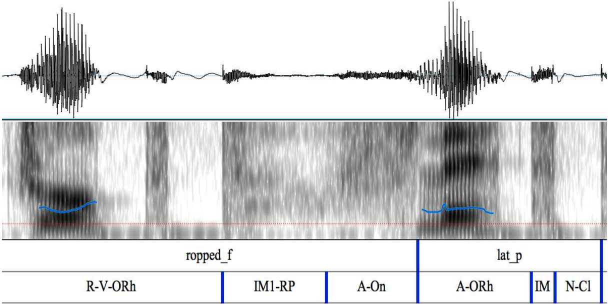
Durational measurements: annotation

- 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 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 precise 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

$$\text{V} \;\;)_\iota(\;\; \text{A} \;\; \text{N}$$

- V-Coda is longer if the verb is infrequent

  27 instances, Int: 0.05, +0.019, t=2,49

- The lower the A-N frequency, the longer the A-Onset

  43 instances, Int: 0.11, +0.05, t=2.42

- The N-Onset is longer if the A-N frequency is low, and if the V-A frequency is higher than the A-N frequency

  87 instances, Int: 0.07, +0.03, t=5.54 // Int: 0.08, + 0.02, t=3.4)

Frequency and duration: some (significant) results II

Syntax B

$$V \quad A \quad)_{\iota}(\quad N$$

- The V-Coda is longer if the V-A frequency is low, or if the verb is infrequent

  28 instances, Int: 0.16, +0.03, t=3.95

  The V-Coda is shorter, if the V-A frequency is larger than the A-N frequency

  28 instances, Int: 0.19, -0.03, t=3.95

- The N-Onset is longer if the Noun is infrequent

  12 instances, Int: 0.09, +0.02, t=2.17

Conclusion duration

All of these results confirm the smooth signal redundancy hypothesis:

$\Rightarrow$ Duration reacts to word and bigram frequency measures in the vicinity of larger prosodic boundaries

Question:

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

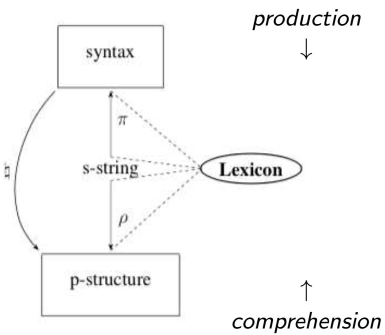
... *performance* knocking at the door of *competence* ...

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)



♮: 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, ...*

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

PHRASING	$(\sigma)_\omega)_\iota$	$(_\iota(\sigma)_\omega$	$(\sigma)_\omega$	...
...	...	...	...	...
LEX_STRESS	prim	prim	prim	...
SEGMENTS	/drɒpt/	/flæt/	/plaŋks/	...
V. INDEX	$\mathbf{S_1}$	$\mathbf{S_2}$	$\mathbf{S_3}$	$\mathbf{S_4}$

... to be further processed (e.g. by postlexical phonology)

- But first: transfer processes to p-structure to create this initial p-diagram

The Transfer of Vocabulary

- Associates morphosyntactic and phonological information on lexical elements
- Via the multidimesional lexicon, which projects them to their respective structures

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

- 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	$('\sigma)_\omega$

P-FORM	[plaŋks]
SEGMENTS	/p l a ŋ k s/
METR. FRAME	$('\sigma)_\omega$

PHRASING	$(\sigma)_\omega$	$(\sigma)_\omega$	$(\sigma)_\omega$	...
...	...	...	...	...
LEX_STRESS	prim	prim	prim	...
SEGMENTS	/drɒpt/	/flæt/	/plaŋks/	...
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	$\mathbf{S_{min}}$	$\mathbf{S_4}$	$\mathbf{S_5}$	$\mathbf{S_{max}}$

- 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**	
drop V	P-FORM	[drɒpt]	FREQ:	high
	SEGMENTS	/d r ɒ p t/	CO-OCURRENCE:	
	METR. FRAME	$('\sigma)_\omega$	drop=flat	high
plank N	P-FORM	[plɑŋks]	FREQ:	low
	SEGMENTS	/p l ɑ ŋ k s/	CO-OCURRENCE:	
	METR. FRAME	$('\sigma)_\omega$	flat=planks	low

Frequency effects in the p-diagram

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

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

- ... passed on to the phonology-phonetics interface ...
- where this information can be transformed into the associated acoustic cues
  → For syntax A:
  - A-Onset longer, as A-N frequency is low
  - N-Onset longer, as A-N frequency is low and V-A frequency is high

Overall framework

S/IP/CP

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

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

$\natural$ ρ

s-form	p-form		meta	
drop V	P-FORM	[drɒpt]	FREQ:	high
	SEGMENTS	/d r ɒ p t/	CO-OCURRENCE:	
	METR. FRAME	$('\sigma)_\omega$	drop=flat	high
plank N	P-FORM	[plɑŋks]	FREQ:	low
	SEGMENTS	/p l ɑ ŋ k s/	CO-OCURRENCE:	
	METR. FRAME	$('\sigma)_\omega$	flat=planks	low

PHRASING	$(\sigma)_\omega)_\iota$	$(_\iota(\sigma)_\omega$	$(\sigma)_\omega$	...
SEGMENTS	/drɒpt/	/flæt/	/plɑŋks/	...
LEX_FREQ	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

- Frequencies affect the strength of prosodic boundaries

→ Allow 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

Thank you!

... questions, comments...?