Why predictability is not predictive without a linguistic theory and a theory of processing. The case of external sandhi

Michael Wagner
McGill University

reporting on joint work with:

Meghan Clayards, Laura Fishman, Jacob Hoover, Oriana Kilbourn-Ceron, Josiane Lachapelle, Morgan Sonderegger, James Tanner

Universität des Saarlandes, July 15 2021

(minor edits July 27 2021)

Outline

Outline

Information theoretic explanations

- Principle of least effort (Zipf, 1949)
- Smooth Signal Hypothesis (Aylett and Turk, 2004): ‘robust information transfer in a potentially noisy environment while conserving effort’
- Uniform Information Density (Levy and Jaeger, 2007; Jaeger, 2010): ‘information density is optimized near the channel capacity’
- ...

Information theoretic explanations

- **Phonetic realization of the message**: Frequent words and contextually predictable words are phonetically shorter than infrequent and less predictable words (when controlling for phonemic content); ...
- **Linguistic structure of the message**: Frequent/predictable words have fewer phonemes on average than less frequent words; optional <u>that</u> is pronounced when encoded message is less predictable (e.g. following constituents); ...

Information theoretic explanations

But to what extend can we derive accurate predictions for linguistic phenomena from such global information-theoretic principles?

- Distribution of prosodic prominence (lexical and phrasal stress) Aylett and Turk (2004) and prosodic boundaries (Turk, 2010) can be explained if they are used as tools to smooth the signal (mostly by affecting duration allocated to given word depending on the information that it carries)
- Currie Hall et al. (2018): 'bias toward accurate transmission of MBUs [Meaning Bearing Units] (e.g. words) leads to more accurate predictions about phonological patterns'

Case study: **External sandhi**

Sandhi

External Sandhi

Phonological process sensitive to property of upcoming (or preceding) word

(1) Tapping: Upcoming word begins with vowel

a. A ca[**t**] **m**eowed!

b. A ca[**r**] **a**ttacked!

(2) Nasal place assimilation: Upcoming place of articulation

a. te[n] fingers

b. te[m] bucks

(3) Liaison in French: Upcoming word begins with vowel

a. des vrai~~**s**~~ **c**opins

b. des vrai[**z**] **a**mis

'real friends'

Information-theoretic account

- **Sandhi as reduction:** Gregory et al. (1999): tapping is a form of reduction, and applies more between cohesive pairs of words; predictability of trigger word increased tapping rate (although mutual information had larger effect) (cf. Jurafsky et al., 2001; Aylett and Turk, 2004; Pluymaekers et al., 2005; Jaeger, 2010, and many others)
  - → sandhi rate should **increase** with predictabilty of trigger word **in reductive processes, but not in non-reductive processes**
- **Sandhi as information about upcoming word:** (Turnbull et al., 2018): nasal assimilation applies more between two words if trigger word is less predictable
  - → sandhi rate should **decrease** with predictability of trigger word

Information-theoretic account

Argument today:

- Irrespective of whether sandhi process is reductive (tapping, nasal assimilation) or non-reductive (liaison), they apply **more** when upcoming word is more predictable
- This not predicted by these two information-theoretic accounts
- It's as predicted by an alternative account in terms of the **locality of production planning**
- The locality of production planning may also explain why the prediction of the information-theoretic account is not observed:
  - less predictable upcoming words are less likely to be anticipated while planning current word
  - speakers may not be able to sufficiently anticipate the upcoming word in precisely the circumstance where providing information about it would be most useful

A methodological point: To test for effects of predictability measures we need control for syntax

Outline

What needs to be explained: Locality and Variability

Locality of Sandhi Phenomena

Sandhi phenomena often only apply **locally**: The two words in question have to be in a certain locality relation to each other.

(Kilbourn-Ceron et al. 2016)

Locality and Variability

How is locality accounted for? Two common approaches in linguistics:

- **Syntactic domains** constrain phonological processes (Cooper and Paccia-Cooper, 1980; Kaisse, 1985; Chen, 1987; Pak, 2008, i.a.).
- **Phonological domains** constrain phonological processes (and are influenced by syntax) (Selkirk, 1986; Kaisse, 1985; Nespor and Vogel, 1986; Odden, 1990; Selkirk, 2011, i.a.)

But **why** do particular processes apply within particular domains?

Locality and Variability

Variability

Sandhi processes are often **variable**. Two types of variability:

(i) **Variability of Application**: Sandhi processes often only apply in a probabilistic way.

(ii) **Variability of Domain**: Sandhi processes often have a variable domain (e.g., locality window widens when speech rate increases, e.g. Kaisse 1985 on fast speech phenomena)

What needs to be explained: Locality and Variability

Current accounts in phonology usually assume the following:

(i) **Variability of Application**: Variable Rules/Variable Constraint Ranking (cf. Anttila, 2002; Coetzee and Kawahara, 2013)

(ii) **Variability of Domain**: Multiple prosodic constituents of a certain type optionally restructure into one constituent of that type or vice-versa.(e.g. Nespor and Vogel, 1986)

But **why** are sandhi processes often variable?

The Locality of Production Planning

Can we make predictions about Locality?

Given the nature of a process, is there anything we can predict about the locality domain in which it is going to apply?

Can we make predictions about variability?

Given the nature of a process, is there anything we can predict about whether it is variable, and the structure of the variability?

The Basic Idea

We need to consider **locality of production planning**.

The Locality of Production Planning

Evidence that phonological planning is very *local*:

- Sternberg 1978: Utterance-initiation-time is sensitive to # of upcoming words, but only to phonological detail (# of σ) of first word
- Levelt (1989): phonological detail is planned over a window roughly the size of a **single prosodic word**

The Locality of Production Planning

Evidence that the size of planning windows is *variable*:

- Lahiri & Wheeldon (1997, 2002) that prosodic size of planning window varies by task
- E.g., the complexity of first prosodic word matters most when planning under time pressure...
- ...while the # of upcoming prosodic words matters most when speakers have more time
- Planning window also varies depending on cognitive load (Swets et al., 2013).

The Locality of Production Planning

Production Planning Hypothesis (PPH)

Sandhi processes are local and variable because the phonological detail relevant to the process may not have been planned yet in time

Tapping

Tapping in American English (Kahn 76, Nespor & Vogel 1986):

Monomorphemic words:
butter, **later** → *pretty much always tapped*

Words within a clause:
If you **meet Ann**, ... → *tapped in fast speech (cf. Kahn 76)*

Across Sentences:
It's **late. I'm** leaving. → *(possible but rare: Kahn 76, Nespor & Vogel 86, ..)*

The Locality of Production Planning

The basic linguistic mechanism:[1]:

[t/d] $\rightarrow$ ɾ / __ V

- Why is tapping **local**? The processing theory: Production planning
  - Planning is local: Process can only apply if upcoming vowel is available
- Why is tapping **variable**?
  - Planning is variable: Scope of planning is affected by many factors

The Locality of Production Planning

Predictions of PPH for Phonological Processes

- Processes Sensitive to **upcoming phonological detail** (e.g. does next word start with vowel?):
  $\rightarrow$ **necessarily local and variable**
- Processes sensitive to **higher-level information**, or **preceding phonological detail** (such as: is there another upcoming word? does preceding word end with vowel):
  $\rightarrow$ **not necessarily local or variable**

Evidence for PPH:

Bailey (2019), Kilbourn-Ceron (2015),Kilbourn-Ceron et al. (2017), Kilbourn-Ceron (2017a), Kilbourn-Ceron (2017c), Kilbourn-Ceron et al. (2020) Kilbourn-Ceron and Goldrick (2020), Lamontagne and Torreira (2017), Tamminga (2018), Tanner et al. (2015), Tanner et al. (2017), Wagner (2011), Wagner (2012)

Outline

Effects of Predictability: Tapping

with Oriana Kilbourn-Ceron  *& Meghan Clayards* 

Experiment/corpus: Kilbourn-Ceron, O., Wagner, M., and Clayards, M. (2017). The effect of production planning locality on external sandhi: A study in /t/. Proceedings CLS, 313–326

Corpus: Kilbourn-Ceron, O., Clayards, M., and Wagner, M. (2020). Predictability modulates pronunciation variants through speech planning effects: A case study on coronal stop realiza- tions. Laboratory Phonology: Journal of the Association for Laboratory Phonology, 11(1).

Effects of Predictability: Tapping

PPH predicts effect of predictability of trigger word:

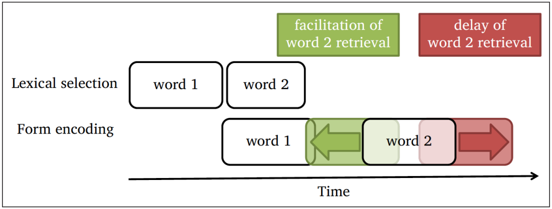


Figure 1: Schema of the time course of select stages of planning a two-word sequence. The vertical dimension represents the distinct stages of linguistic planning, lexical selection and form encoding. The boxes represent the duration of processing at each stage for each word. If conditions facilitate retrieval of the phonological code *word 2*, form encoding can begin sooner, and may overlap with form encoding of *word 1* (green scenario), while a delay could prevent interaction between *word 1* and *word 2* (red scenario).

Tapping: Corpus Data

- We looked at Buckeye Corpus (Pitt et al., 2007). to look for effect predictability measures
- 11863 tokens with word-findal /t/ or /d/ followed by a vowel-initial word (46.24% were transcribed as flaps).
- Excluded: words followed by disfluency (18.26% of tokens)
- Word frequencies were retrieved from SUBTLEX-US, a database of word frequencies based on film and television subtitles (Brysbaert and New, 2009)

Tapping: Corpus Data

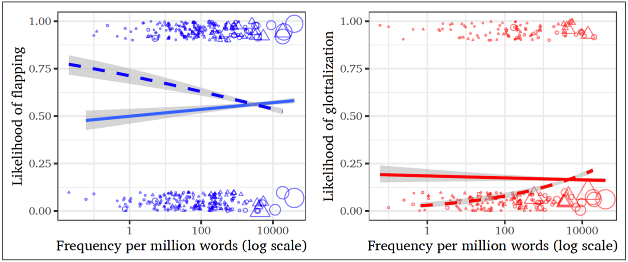


Figure 2: Relationship between SUBTLEX-US word frequency (per million words) and proportion of tokens transcribed as flaps [dx] (left panel, blue), glottal stops [tq] (right panel, red) in the Buckeye corpus. Solid lines show trigger word frequency, dashed lines show target word frequency, with shading showing 95% confidence intervals of a linear smooth (GLM, logit-link). Bubbles (trigger word frequency) and triangles (target word frequency) show density of observations, with size reflecting the number of observations at a given value on the x-axis.

Tapping & Glottalization: Conditional probability

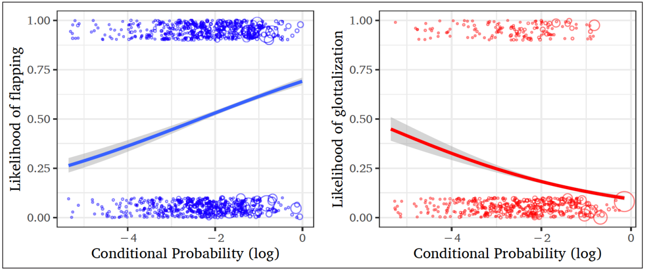


Figure 3: Relationship between Conditional Probability (of trigger word given target word) and proportion of tokens transcribed as flaps [dx] (blue, left panel) or glottal stops [tq] (red, right panel) in the Buckeye corpus. Solid lines and shading are linear smooths (GLM, logit-link) with 95% confidence intervals. Bubbles show density of observations, with size reflecting the number of observations at a given value on the x-axis.

Tapping & Glottalization: Frequency Effects

- Higher **Target Word Frequency** $\rightarrow$ less tapping, more glottalization
- Higher **Trigger Word Frequency** $\rightarrow$ more tapping, no effect on glottalization
- Higher **Trigger Word Frequency** $\rightarrow$ more tapping, less glottalization

Why do these two reductive processes pattern differently?

Tapping vs. Glottalization

- Tapping only possible if following word begins with vowel
- Glottalization does not require information about next word:
  - Tapping: [t/d] $\rightarrow$ ɾ / __ V
  - Glottalization: [t/d] $\rightarrow$ ʔ/ __ #
- $\rightarrow$ PPH predicts predictability of trigger word will facilitate tapping
- glottalization rate decreases since tapping rate increases

(see Seyfarth and Garellek (2020) for other contextual effects on glottalization rate)

Tapping vs. Glottalization

Why negative effect of Target Word Frequency on tapping rate?

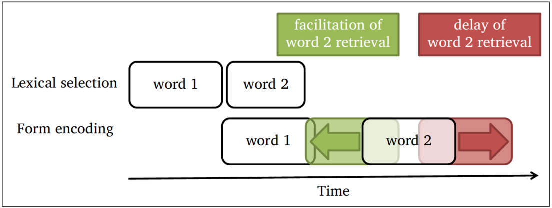


Figure 1: Schema of the time course of select stages of planning a two-word sequence. The vertical dimension represents the distinct stages of linguistic planning, lexical selection and form encoding. The boxes represent the duration of processing at each stage for each word. If conditions facilitate retrieval of the phonological code *word 2*, form encoding can begin sooner, and may overlap with form encoding of *word 1* (green scenario), while a delay could prevent interaction between *word 1* and *word 2* (red scenario).

Word1 frequency increase means it may be planned earlier relative to planning of word2, unless the two are planned as single word (e.g. determiner + noun)

Can information-theoretic account explain the differences between the processes?

- Effect of trigger word predictability compatible with sandhi-as-reduction account (Gregory et al., 1999), but effect of target word and differences to glottalization not predicted
- Effect of trigger word predictability unexpected under sandhi-as-information-about-upcoming-word account (Turnbull et al., 2018) (but that account could be combined with sandhi-as-reduction effects)

Outline

- Nasal assimilation encodes information about the upcoming word
- It is reductive in that assimilation facilitates compression of two-word sequence, and categorical assimilation at least involves loss of information
- Turnbull et al. (2018) 'More assimilation should be observed for more contextually predictable target words, while less assimilation should be observed for contextually more predictable trigger words.'
- The PPH predicts trigger-word predictability should facilitate assimilation irrespective of the adaptiveness from a global information-theoretic point of view—nasal assimilation should be more likely if following word is more predictable

Turnbull et al. (2018): findings for the effect of trigger word predictability:

- Trigger word predictability, categorical assimilation: No effect
- Trigger word predictability, gradient assimilation: Decrease in conditional probability of the trigger word given the target word increases assimilation rate
- Within-word predictability effect for categorical assimilation: If an upcoming word begins with a less frequent phoneme (essentially: smaller lexical cohort size), there is more nasal assimilation (see Appendix for discussion)

These results seem problematic for the PPH!

Syntax and predictability

Syntax and frequency/predictability are confounded:

- In a head-initial language like English, the first word of a syntactic constituent is often a function word and hence high frequency
- Nasal assimilation might be blocked by a juncture precisely in those cases in which a very <u>frequent</u> word follows
- The conditional predictability of of the trigger word given the target word may also depend on syntax (within a constituency upcoming words may be more predictable)

Syntax and predictability

- We extracted two-word sequence with potential for nasal assimilation, following Turnbull et al. (2018)
- Two annotations (MW and one undergraduate RA) hand-annotation of syntactic juncture (0 = word or lower; 1 = XP; 2 = Clausal boundary)

Most frequent words at clausal juncture:

w2	n	Proportion
but	95	57.20
because	46	27.70
before	4	2.40
basically	3	1.80

Table: words by syntax

- The four most common words are function words
- They account for almost 90% of all words following a strong boundary

Syntax and predictability

Assimilation rate by syntax:

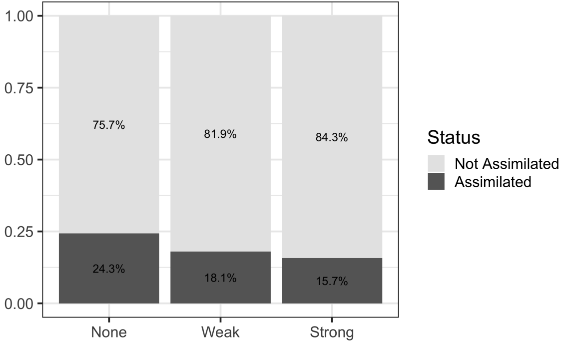


Figure: The assimilation rate depending on the following syntactic juncture.

Syntax and predictability

Relationship between word frequency and syntax:

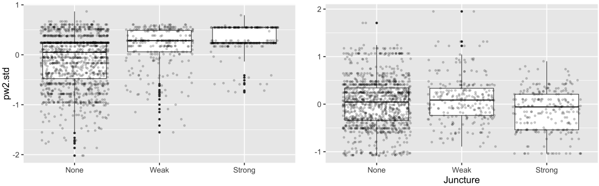


Figure: Frequency (left) and predictability (right) of trigger word depending on preceding syntactic juncture

Syntax and predictability

	pw1	pw2w1
(Intercept)	$-0.00(0.02)$	$0.01(0.02)$
Strong.vs.Other	$-0.37(0.04)^{***}$	$0.14(0.05)^{**}$
Weak.vs.None	$-0.20(0.03)^{***}$	$-0.01(0.03)$

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Testing effect of syntactic position on log frequency of trigger word (pw2) and the log of the conditional probability of the trigger word given the target word (pw2w1).

- Syntactic junctures correlate with predictability measures
- This may not come as a surprise, e.g. Turk and Shattuck-Hufnagel (2007) argued that prosodic junctures, traditionally viewed as reflexes of syntactic structure, actually encode points of low predictability
- Apparent effect of frequency/predictability might be due to syntax, or vice versa

Effect of word frequency (SUBTLEX-US)

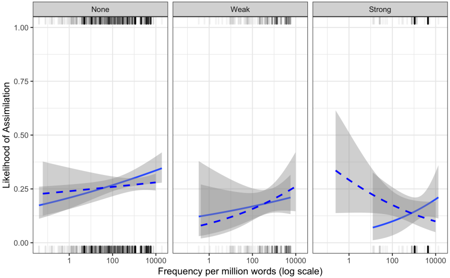


Figure: Effect of frequency of the trigger word (solid line) and the target word (dashed line) on assimilation, depending on syntax

Effect of trigger-word predictability (COCA corpus Davies, 2011)

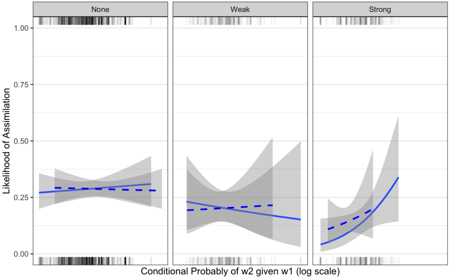


Figure: Effect of the conditional probability of the second word given the first (solid line) and the first word given the second (dashed line).

Effect of predictability: XLnet

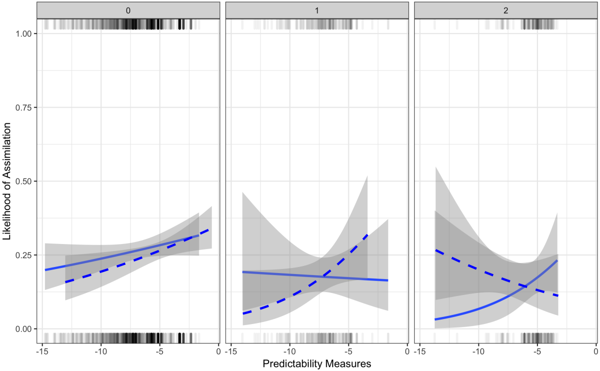


Figure: Predictability measures from XLnet. The solid line shows the probability of the trigger word given all other words in the sentence; the dashed line shows an estimate of the log probability of the target word given all other words in the sentence.

Thanks to Jacob Hoover for extracting these estimates

Effect of predictability: measures in Turnbull et al. 2018t

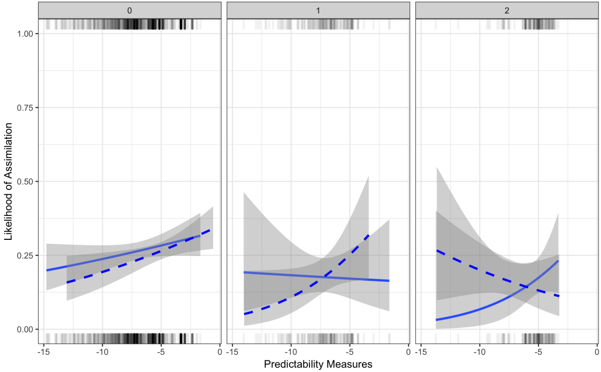


Figure: Predictability measures from Turnbull et al. 2018. Effect of the conditional probability of the second word given the first (solid line) and the first word given the second (dashed line).

	P	XLnet	T.2018
(Intercept)	$-1.96(0.33)^{***}$	$-1.89(0.32)^{***}$	$-2.16(0.35)^{***}$
Strong.vs.Other	$0.40(0.63)$	$0.33(0.55)$	$0.41(0.61)$
Weak.vs.None	$0.79(0.26)^{**}$	$0.84(0.23)^{***}$	$0.73(0.25)^{**}$
g.vs.other	$-0.51(0.39)$	$-0.58(0.38)$	$-0.48(0.33)$
b.vs.kp	$-0.78(0.27)^{**}$	$-0.80(0.27)^{**}$	$-0.76(0.25)^{**}$
p.vs.k	$0.01(0.26)$	$-0.17(0.26)$	$-0.02(0.23)$
phone_duration.std	$-1.77(0.32)^{***}$	$-1.59(0.29)^{***}$	$-1.68(0.28)^{***}$
pw1.std	$0.14(0.37)$	$-0.22(0.25)$	
pw2.std	$0.42(0.38)$	$-0.15(0.37)$	
pw2w1.std	$0.27(0.28)$		
Strong.vs.Other:pw2w1.std	$-1.44(0.80)$		
Weak.vs.None:pw2w1.std	$0.11(0.46)$		
xw1.std		$0.29(0.23)$	
xw2.std		$0.73(0.31)^{*}$	
Strong.vs.Other:xw2.std		$-0.28(0.82)$	
Weak.vs.None:xw2.std		$0.23(0.45)$	
tw1.std			$0.46(0.49)$
tw2.std			$-0.64(0.42)$
tw1w2.std			$-0.29(0.50)$
tw2w1.std			$1.30(0.54)^{*}$
Strong.vs.Other:tw2w1.std			$-1.94(1.38)$
Weak.vs.None:tw2w1.std			$0.38(0.48)$

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table: Model including both normalized of the log of predictability and syntactic predictors, as well as phone duration as a proxy for prosody.

Summary of look at nasal assimilation

- Syntax needs to be considered when testing for predictability effects
- At least once syntax is taken into account, it seems that trigger word predictability increases assimilation rate, as predicted by PPH
- Important limitations:
  - We only looked at categorical assimilation so far, but the effect in the opposite direction in Turnbull et al. (2018) was only observed for gradient assimilation
  - The model results are volatile due to colinearity between predictors, and we used less sophisticated statistical methods compared to Turnbull et al. (2018)
  - See appendix for within-word-predictability (cohort-size)

Summary of look at nasal assimilation

- Are syntactic effects confounded with predictabilty effects, or *are* syntactic effects really predictability effects?
- Either way, the direction of the effect is as predicted by PPH: Across syntactic boundaries the upcoming word is less predictable, and sandhi processes are less likely to apply
- Across syntactic junctures...
  - ... ing/in choice is less influenced by following segment (Wagner, 2011, 2012)
  - ... tapping is less likely to cue an upcoming vowel (Kilbourn-Ceron et al., 2017)
  - ... t/d deletion is less influenced by following segment (Tamminga, 2018)

Outline

Effects of predictability in non-reductive processes: Liaison

with Oriana Kilbourn-Ceron  *& Josiane Lachapelle* 

Corpus Study: Kilbourn-Ceron, Oriana (2016). Speech production planning affects variability in connected speech. Proceedings of AMP, USC

Experimental study: Wagner, M., Lachapelle, J., and Kilbourn-Ceron, O. Liaison and production planning. Poster presentation and the 17th conference on Laboratory Phonology at UBC.

Liaison: Latent consonant appears before vowel initial word

Un peti[t] ami

cf. un peti[ ] chapeau

Des vrai[z] amis

cf. des vrai[ ] chatons

slide Oriana Kilbourn-Ceron

Liaison: Predictions of information-theoretic account

Turnbull et al. (2018):

- 'conserve cost when message predictability is high'
- ...and 'additional material increasing signal specificity and redundancy is more likely to be invested when message predictability is low'

PPH makes **opposite** prediction: more predictable trigger word $\rightarrow$ more liaison

Liaison: Production experiment

- Another factor affecting planning scope: **word length**
- If word1 is long, then it is less likely that word2 will be planned at the same time (Miozzo and Caramazza, i.a.—but: Griffin)
- Since only the *beginning* of word2 is relevant (*does it start with a vowel?*), its overall length might be less relevant
- Also manipulated: speech rate, repetition, word frequencies, conditional probability of upcoming word, syntax (adjective-noun vs. noun-adjective contexts)

Liaison: Production experiment

- 20 speakers (Québecois and European French)
- Each recorded on 80 sentences and their (back-to-back) repetitions
- We asked to talk as naturally as possible, as if in a conversation
- Data were annotated for liaison and analyzed using ME logistic regression

Liaison: Production experiment

(4) Adjective-Noun

a. Low conditional probability; short word 1; short word 2:

Elle discute avec les **derniers élèves**.
she discusses with the last students
'She is talking with the latest students.'

b. High conditional probability, short word1; short word2:

Vous regrettez vos **dernières années**.
you regret your last years
'You regret the previous years.'

(5) Noun-Adjective)

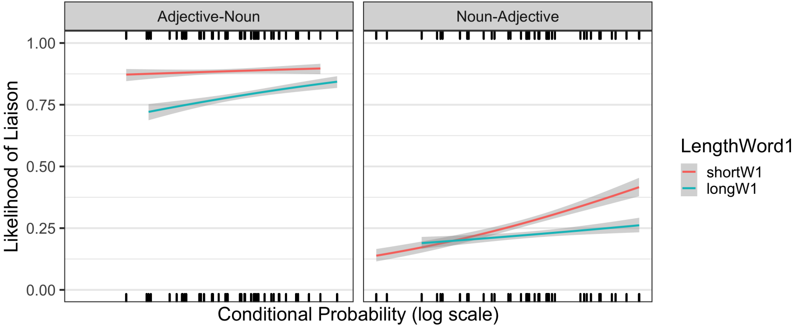
a. Low conditional probability; short word 1; long word 2:

Ils construisent des **douches intérieures**.
they construct of douches interior
'They are constructing interior showers.'

b. High conditional probability; short word 1; long word 2:

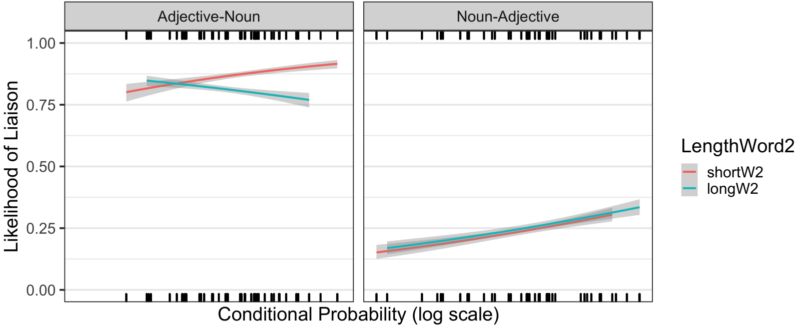
Mathilde regarde ses **dessins animés**.
Mathilde watches her drawing animated
'Mathilde is watching her cartoons.'

Liaison: Production experiment



Plot of the effect of conditional probability, syntax, and length of word1

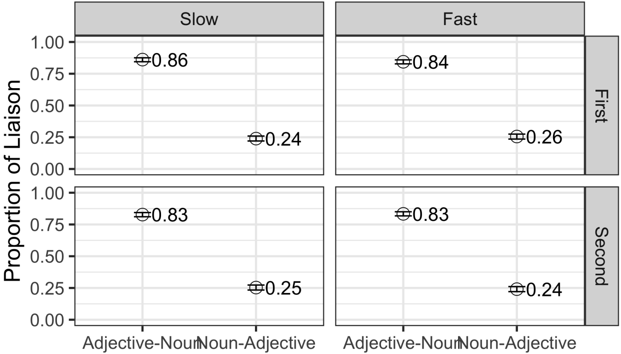
Liaison: Production experiment



Plot of the effect of conditional probability, syntax, and length of word2

Liaison: Discussion

Effects of speech rate and repetition:



Liaison: Discussion

- Conditional probability of word2 given word 1 increases liaison rate
- Length of Word1 matters when liaison rate not close to ceiling or floor
- Syntactic proximity matters
- Speech rate does not matter (already observed in Kaisse)
- Repetition does not matter

- Lack of speech rate (cf.Kaisse (1985)) or repetition effect seems surprising based on PPH, and contrasts results from other cross-word phonological processes, e.g. tapping Kilbourn-Ceron (2017b,c)
- Bybee (2001) and Côté (2013) attribute frequency effects in liaison to the storage of larger sized units. This could explain absence of these effects.
- Maybe storage fossilizes effects of the PPH in (at least partially) lexicalized processes like liaison

Liaison: Conclusion

- Results are in line with predictions of PPH
- Results conflict with information-theoretic rationale
- But the locality of production planning may also explain why the information-theoretic expectation is not observed: Even if it would have been beneficial for transmission in the channel, you cannot cue that an upcoming word starts with a vowel if you don't know its phonological shape yet (and less predictable words are less likely to be planned early relative to preceding word)

Outline

Toward a predictive theory of locality & variability in phonology

- Chen (1987, 2000): Locality of Tone sandhi in Taiwanese (also Xiamen) is constrained by syntax, but often contradicts prosody
- ...other types of tone sandhi, e.g. Mandarin T3 sandhi, seem to be much more variable and constrained by surface prosody

Why do these processes differ in their locality and variability?

Toward a predictive theory of locality & variability in phonology

- Taiwanese
  - every non-final word within a domains undergoes tone sandhi;
  - The following tone is irrelevant in determining which sandhi tone it shifts to.
  - Crucially, the only information relevant is whether a word is coming up within the same syntactic domain.
- Mandarin T3 sandhi
  - Which sandhi tone you shift to depends on phonological identity of following tone
  - $\rightarrow$ the phonology of the following word has to have been planned out for T3 sandhi to apply
  - The PPH predicts the process to be local and variable.

(Wagner, 2012)

More predictions: Influence of prior vs. upcoming information in vowel coalescence (Lamontagne and Torreira, 2017)

Conclusions

Can we make predictions about Locality?

Maybe yes: When a process relies on phonological information about an upcoming word, it should necessarily be local; when it depends on phonological information about a previous word, or on higher level information, it does not need to be local.

Can we make predictions about variability?

Maybe yes: If a process relies on phonological information contained in an upcoming word, it necessarily has to be variable, but not if it relies on information from preceding word.

Conclusions

Is the information-theoretic perspective wrong?

- Arguably, the data doesn't necessarily speak against the global information-theoretic rationale
- Rather, the data may reveal limitations of what speakers can actually anticipate and calculate online when they plan and realize the message
- Still, this means that when it comes to understanding trigger-word predictability effects on sandhi, the information-theoretic view may not explain much of the variability

Conclusions

However: the data suggests that not every predictability effect subserves a global information-theoretic rationale (e.g., contributing to a smooth signal):

- The predictions of the PPH are derived from the assumed mechanism (phonological rule application) and known constraints on production planning (the processing mechanism)
- The predicted (and observed observed) predictability effects do not seem to optimize transmission in the channel according to a global rationale
- It could still be, of course, be that sandhi patterns that optimize the signal are more likely to be grammaticalized, or to remain productive longer
- Could it be that liaison is so different from tapping precisely because it is not reductive? (liaison seems to have been lexicalized, hence no speech rate effect)

Conclusions

- Accessibility/repetition effects
- The predicted (and observed observed) predictability effects do not seem to optimize transmission in the channel according to a global rationale
- It could still be, of course, be that sandhi patterns that optimize the signal are more likely to be grammaticalized, or to remain productive longer
- Could it be that liaison is so different from tapping precisely because it is not reductive? (liaison seems to have been lexicalized, hence no speech rate effect)

Thanks!

Co-authors:

Oriana Kilbourn-Ceron  *Meghan Clayards*  *Josiane Lachapelle* 

James Tanner  *Morgan Sonderegger* 

Thanks to **SSHRC, NSERC**, and to the members of the **prosody.lab**, **speech.learning.lab**, and **mlml lab** at McGill for their help.

Thanks also to Michael McAuliffe for helping us extract data from Buckey Corpus.

Within-word predictability (cohort size) revisited

Cohort effects: There are four segments, ordered from smallest cohort [g] (lowest ‘within-word-predicability) to biggest cohort [k] (highest within-word-predictability):

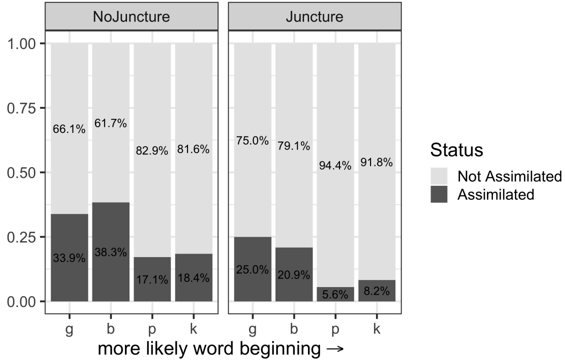


Figure: The assimilation rate depending on the following phoneme and the presence/absence of a syntactic juncture.

Within-word predictability (cohort size) revisited

- There is a lower assimilation rate when there is a syntactic juncture.
- The assimilation rate is higher when the following stop is voiced compared to when it is voiceless
- There's only four segments, but given the plot it seems that there is an effect of voicing on assimilation
- It doesn't seem that there is an effect of cohort-size

Within-word predictability (cohort size) revisited

Two potential reasons why voicing might affect the (perceived) assimilation rate:

- Articulatory: Place assimilation with a following voiced stop make it easier to maintain voicing throughout, and shared voicing may make it easier to share place gesture
- Perceptual: Dilley and Pitt (2007) report that whether assimilation is transcribed depends closure duration. Maybe it is simply harder to hear the separate place of articulation of the nasal if the following stop has a shorter closure duration, like voiced consonants do.
- A perceptual explanation is also compatible with the fact that within-word predictability was not found to contribute to explaining acoustic correlates of partial assimilation in Turnbull et al. (2018).

Within-word predictability (cohort size) revisited

Does information-theoretic account actually predict that cohort-size should decrease assimilation rate?

- Smaller cohort size: Upcoming **segment** is **less** predictable
- Smaller cohort size: Upcoming **word** is **more** predictable once initial segment is identified

$\rightarrow$ If predictability effects are about predicting meaningful units, as argued in Currie Hall et al. (2018), then the prediction could have been be that larger cohort will require advance warning, not the smaller cohort

Accent placement

Since I made reference to accent placement in my abstract:

- Smooth signal hypothesis proposes that prosodic prominence is tool to manage information density
- This would predict that less predictable/accessible words more likely to carry accent (e.g., if word more accessible/less surprising → less likely to carry accent)
- Wagner and Klassen (2015): After confounds with focus are controlled for, accessibility and repetition do not contribute to predict accent placement
- Klassen and Wagner (2017): After confounds with focus are controlled for, accessibility, predictability, and repetition do not contribute to predict accent placement
- There are effects of repetition/accessibility/predictability, but they are not observed on the target word but on words preceding target word
- There are corners of the syntax where accent placement is affected by predictability, but these are mediated by information structure (e.g. stress in intransitives)

Anttila, A. (2002). Variation and phonological theory. In The handbook of language variation and change, pages 206–243. Wiley-Blackwell, Oxford.

Aylett, M. and Turk, A. (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.

Bailey, G. (2019). Ki(ng) in the north: effects of duration, boundary and pause on post-nasal [g]-presence. Laboratory Phonology.

Brysbaert, M. and New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. Behavior research methods, 41(4):977–990.

Bybee, J. (2001). Frequency effects on French liaison. In Bybee, J. and Hopper, P., editors, Frequency and the emergence of linguistic structure, pages 337–359. John Benjamins.

Chen, M. (2000). Tone sandhi: Patterns across Chinese dialects. Cambridge Univeristy Press.

Chen, M. Y. (1987). The syntax of Xiamen tone sandhi. Phonology Yearbook, 4:109–49.

Coetzee, A. W. and Kawahara, S. (2013). Frequency biases in phonological variation. Natural Language & Linguistic Theory, 31(1):47–89.

Cooper, W. E. and Paccia-Cooper, J. (1980). Syntax and Speech. Harvard University Press, Cambridge, Mass.

Côté, M.-H. (2013). Understanding cohesion in french liaison. Language Sciences, 39:156–166.

Currie Hall, K., Hume, E., Jaeger, T. F., and Wedel, A. (2018). The role of predictability in shaping phonological patterns. Linguistics Vanguard.

Davies, M. (2011). N-grams data from the corpus of contemporary american english (coca). Downloaded from http://www.ngrams.info on August 16, 2020.

Dilley, L. C. and Pitt, M. A. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. The Journal of the Acoustical Society of America, 122(4):2340–2353.

Gregory, M., Raymond, W., Bell, A., Fosler-Lussier, E., and Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. In Proceedings of the Chicago Linguistic Society, volume 35, pages 151–166.

Griffin, Z. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. Psychonomic Bulletin and Review, 10:603–609.

Jaeger, T. (2010). Redundancy and Reduction: Speakers Manage Syntactic Information Density. Cognitive Psychology, 61(1):23–62.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. and Hopper, P., editors, Frequency in the Emergence of Linguistic Structure, pages 229–254. John Benjamins, Amsterdam.

Kaisse, E. M. (1985). Connected Speech. The interaction between syntax and phonology. Academic Press, Orlando, Flo.

Kilbourn-Ceron, O. (2015). The influence of prosodic context on high vowel devoicing in spontaneous Japanese. In Proceedings of the 18th International Congress of Phonetic Sciences (ICPHS) in Glasgow.

Kilbourn-Ceron, O. (2017a). Speech production planning affects phonological variability: a case study in French liaison. In Proceedings of the Annual Meeting on Phonology 2016, October 21-23, 2016. University of Southern California, Los Angeles, CA.

Kilbourn-Ceron, O. (2017b). Speech production planning affects variability in connected speech. Proceedings of the Annual Meeting on Phonology (AMP), USC.

Kilbourn-Ceron, O. (2017c). Speech production planning affects variation in external sandhi. PhD thesis, McGill University.

Kilbourn-Ceron, O., Clayards, M., and Wagner, M. (2020). Predictability modulates pronunciation variants through speech planning effects: A case study on coronal stop realizations. Laboratory Phonology: Journal of the Association for Laboratory Phonology, 11(1).

Kilbourn-Ceron, O. and Goldrick, M. (2020). Larger phonological planning windows trigger variation in word-final consonants. Talk at LabPhon 17, University of British Columbia.

Kilbourn-Ceron, O., Wagner, M., and Clayards, M. (2017). The effect of production planning locality on external sandhi: A study in /t/. In Proceedings of the 52nd Annual Meeting of the Chicago Linguistic Society, pages 313–326.

Klassen, J. and Wagner, M. (2017). Prosodic prominence shifts are anaphoric. Journal of Memory and Language, 92:305–326.

Lamontagne, J. and Torreira, F. (2017). Production planning effects in sandhi: A corpus study using automated classification. Talk presented at New Ways of Analyzing Variation 46, Madison, Wisconsin.

Levy, R. and Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schlökopf, J. P. and Hoffman, T., editors, Advances in neural information processing systems (NIPS), volume 19, pages 849–856.

Miozzo, M. and Caramazza, A. (2003). When more is less: a counterintuitive effect of distractor frequency in the picture-word interference paradigm. Journal of Experimental Psychology: General, 132(2):228.

Nespor, M. and Vogel, I. (1986). Prosodic Phonology. Foris, Dordrecht.

Odden, D. (1990). Syntax, lexical rules and postlexical rules in Kimatuumbi. In Inkelas, S. and Zec, D., editors, The phonology-syntax Connection, pages 259–278. CSLI and CUP.

Pak, M. (2008). The postsyntactic derivation and its phonological reflexes. PhD thesis, University of Pennsylvania.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., and Fosler-Lussier, E. (2007). Buckeye corpus of conversational speech (2nd release). Columbus, OH: Department of Psychology, Ohio State University.

Pluymaekers, M., Ernestus, M., and Baayen, R. (2005). Articulatory planning is continuous and sensitive to informational redundancy. Phonetica, 62(2-4):146–159.

Selkirk, E. (1986). On derived domains in sentence phonology. Phonology Yearbook, 3:371–405.

Selkirk, E. (2011). The syntax-phonology interface. In Goldsmith, J., Riggle, J., and Yu, A., editors, The Handbook of Phonological Theory, pages 435–484. Blackwell, Oxford, 2nd edition.

Seyfarth, S. and Garellek, M. (2020). Physical and phonological causes of coda/t/glottalization in the mainstream american english of central ohio. Laboratory Phonology: Journal of the Association for Laboratory Phonology, 11(1).

Swets, B., Jacovina, M. E., and Gerrig, R. J. (2013). Effects of conversational pressures on speech planning. Discourse Processes, 50(1):23–51.

Tamminga, M. (2018). Modulation of the following segment effect on english coronal stop deletion by syntactic boundaries. Glossa.

Tanner, J., Sonderegger, M., and Wagner, M. (2015). Production planning and coronal stop deletion in spontaneous speech. In Proceedings of the 18th International Congress of Phonetic Sciences (ICPHS) in Glasgow.

Tanner, J., Sonderegger, M., and Wagner, M. (2017). Production planning and coronal stop deletion in spontaneous speech. Laboratory Phonology, 8 (1): 15:1–39.

Turk, A. (2010). How does speech +ming work? Lecture Notes, SPSASSD, Brasil.

Turk, A. and Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. Journal of Phonetics, 35(4):445–472.

Turnbull, R., Seyfarth, S., Hume, E., and Jaeger, T. F. (2018). Nasal place assimilation trades off inferrability of both target and trigger words. Laboratory Phonology: Journal of the Association for Laboratory Phonology, 9(1).

Wagner, M. (2011). Production-planning constraints on allomorphy. Canadian Acoustics, 39(3):160–161.

Wagner, M. (2012). Locality in phonology and production planning. In Proceedings of Phonology in the 21 Century: Papers in Honour of Glyne Piggott, Montréal. McGill Working Papers.

Wagner, M. and Klassen, J. (2015). Accessibility is no alternative to alternatives. Language, Cognition, and Neuroscience, 30:212–233.

Zipf, G. K. (1949). Human Behavior and the Principle of Least Effort. Addison-Wesley.