Japanese speakers can infer specific sub-lexicons using phonotactic cues

Shigeto Kawahara & Gakuji Kumagai

Phonology Festa

2024/03/07

Background

- Phonotactic restrictions do not usually hold uniformly across the entire lexicon.

- A famous example: Japanese has several “lexical strata”, which are sensitive to phonotactic restrictions to different degrees (Ito & Mester 1995, 1999). E.g. voiced obstruent geminates are not allowed, except in loanwords.

- A question that is addressed in some recent studies: how fined-grained can these sub-lexicons be? (e.g. Gouskova et al. 2015)

Two broad hypotheses

- The “classic view”: Languages only allow “general” divisions (perhaps with independent motivations, historical and/or orthographic).
  - Japanese: native, Sino-Japanese, recent loanwords, mimetics (Ito & Mester 1999).
  - English: Germanic vs. Latinate (Chomsky & Halle 1968).
- The emerging view: Sublexicons can be as specific as “a set of roots to which a particular affix is attached to” (Gouskova et al. 2015; see also the body of references on “co-phonology (by phase)”: Sande 2020).

The sublexal phonology/co-phonology

- “learning lexically specific morphological and phonological rules involves separating the lexicon into sublexicons. Phonological generalizations about the application of such rules are encoded in part as phonotactic grammars learned over sublexicons.” (Gouskova & Ahn to appear, page 6).

- “Cophonology Theory is motivated by the non-uniformity of the phonological grammar of single languages. In this theory, every language contains multiple phonological sub-grammars which apply in different morphosyntactic environments. To date, cophonologies have been indexed to part of speech (Anttila 2002; Smith 2011) and to specific morphological constructions (Orgun 1996; Inkelas 1998; Inkelas and Zoll 2005, 2007)” (Sande 2020)

Our hunch

- Some sound sequences in Japanese are over-represented -- or restricted to -- a very specific word class (i.e. a sub-lexicon).
  - Geminate [rr]s are generally not allowed in Japanese, but they often appear on “Italian restaurant menus”.
  - Geminate [hh]s are also very rare, except in Dutch/German names.
  - Snack names often contain singleton [p]s.

Expanding on the last point

- In July 2023, I [Shigeto] published this book, “Why snack names contain a lot of [p]s?”.
- This is based on an actual lecture that I gave to elementary school students.
- Since I had never taught linguistics to elementary school students, prior to the event, I had asked the participants to send me “whatever questions they have about languages”.

Expanding on the last point

- One of the questions was, “I feel like [p] is often used in snack names, like *papiko* and *poihuru*. Why is that?”
- I thought that this was a fantastic question. And she provided many actual examples!
- So I provided some possible explanations (not directly relevant here).
- The publisher thought that this is the catchiest question, and it became the main title of the book.

よくおかしの商品名に
「パ」とか「ピ」が付けられているけどなんで？

川原　じゃあ、次の質問。「よくお菓子の商品名にパとかピが付けられているけどなんで？」。わこ、素晴らしい質問だね。わこがあげてくれた例は「パピコ」「パイの実」「チョコパイ」「アポロ」「ポイフル」ですね。

Expanding on the last point

- But a few months later after the publication of this book, I realized that “why X” presupposes that “X is correct”.
- But I did not actually test whether X is correct in that very book.
- As a linguist, I decided that it is my responsibility to test X, the presupposition.
- And this motivation converged with the theoretical question discussed at the outset.

Three hunches tested

1. Singleton [p]s are overrepresented in snack names.

2. Geminate [rr]s almost exclusively appear in Italian loanwords.

3. Geminate [hh]s appear often in German names, but not anywhere else.

Experiment 1

- Japanese speakers were presented with a target word containing the phonotactic cue (e.g. [**p**ariko]) and a control word not containing that phonotactic cue (e.g. [**t**ariko]) .

- They were asked to choose whether that name is suitable for the target category (e.g. a European snack) or a control category (e.g. a cosmetic brand).

- We presented one item per trial, not in a pair (i.e. a forced choice task, not a 2AFC).

	Target	Control
[p]=snack	[paɾiko]	[taɾiko]
	[pasomi]	[tasomi]
	[penaɾo]	[tenaɾo]
	[posine]	[tosone]
	[pamoɾe]	[kamoɾe]
	[paɾase]	[kaɾase]
	[pesemo]	[tesemo]
	[ponoɾu]	[koniɾu]
[ɾɾ]=Italian	[metoɾɾa]	[metobba]
	[makoɾɾo]	[makobbo]
	[nesoɾɾa]	[nesodda]
	[nokiɾɾe]	[nokidde]
	[temeɾɾo]	[temeddo]
	[tamiɾɾa]	[tamigga]
	[kanoɾɾi]	[kanoggi]
	[tonoɾɾe]	[tonogge]
[hh]=German	[bohho]	[boppo]
	[kuɸɸu]	[kuppu]
	[gehhe]	[geppe]
	[gahha]	[gatta]
	[ɾehhe]	[rette]
	[biççi]	[bikki]
	[nahha]	[nakka]

The stimuli

- Target=[p]-initial words; control=[t/k]-initial words.
- Target items contain [rr]; control items contain a voiced obstruent geminate.
- Target items contain [hh]; control items contain a voiceless obstruent geminate.

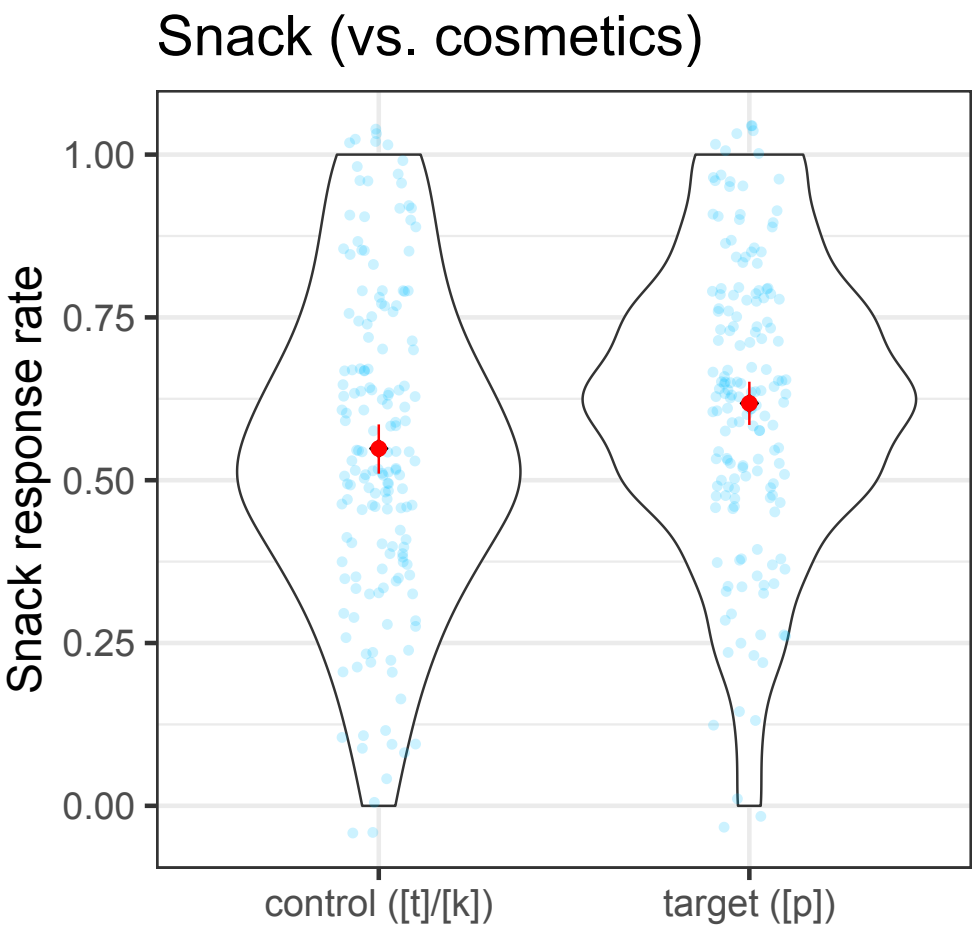
	Target	Control
[p]=snack	[pariko]	[tariko]
	[pasomi]	[tasomi]
	[penaro]	[tenaro]
	[posine]	[tosone]
	[pamore]	[kamore]
	[parase]	[karase]
	[pesemo]	[tesemo]
	[ponoru]	[koniru]
[rr]=Italian	[metorra]	[metobba]
	[makorro]	[makobbo]
	[nesorra]	[nesodda]
	[nokirre]	[nokidde]
	[temerro]	[temeddo]
	[tamirra]	[tamigga]
	[kanorri]	[kanoggi]
	[tonorre]	[tonogge]
[hh]=German	[bohho]	[boppo]
	[kuɸɸu]	[kuppu]
	[gehhe]	[geppe]
	[gahha]	[gatta]
	[rehhe]	[rette]
	[biççi]	[bikki]
	[nahha]	[nakka]

The task

- Snack names or cosmetic names? (they are both written in the *katakana* orthography)
- Italian restaurant names or French restaurant names?
- German names or English names?

Experiment details

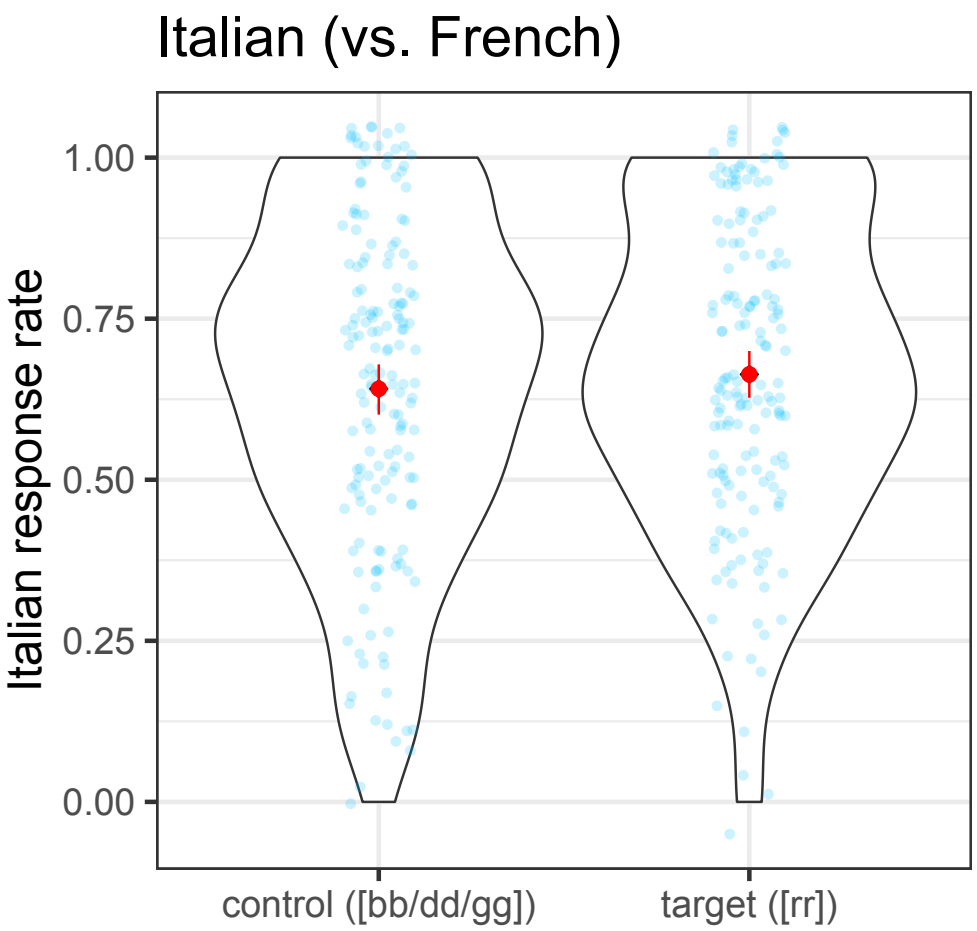
- Data from 162 participants were collected.
- Administered online using SurveyMonkey.
- The stimuli were presented in the *katakana* orthography.
- The data was analyzed using a Bayesian mixed effect logistic regression model. Full details available on an osf repository.
- We present the posterior probability of the slope coefficient (β1) to be positive.



Results

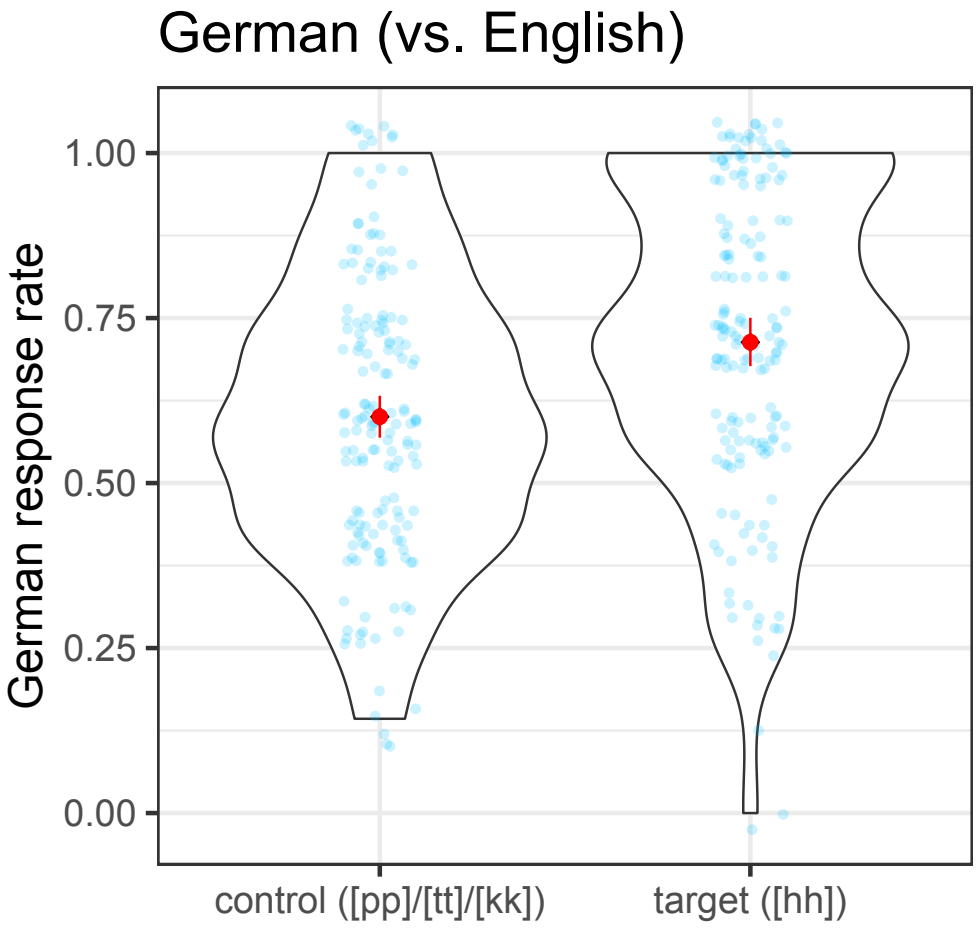
- Those items that contain [p] were more likely judged to be snack names than the control items.

- $p(\beta 1>0) = 0.95$.



Results

- Those items that contain [rr] were more likely judged to be Italian names than the control items, but not by much.
- $p(\beta 1>0) = 0.82$.
- So not as clear as the association between [p] and snacks.



Results

- Those items that contain [hh] were more likely judged to be German names than the control items.

- $p(\beta 1>0) = 0.95$.

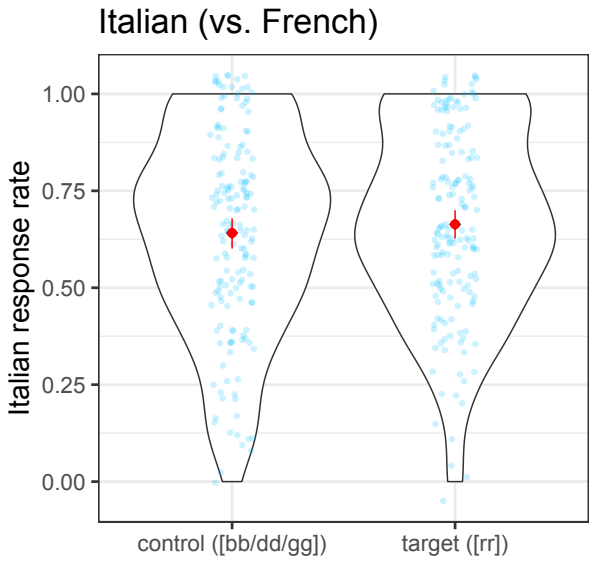
Discussion

- Japanese speakers can associate nonce words with a particular phonotactic cue to a very specific lexical class.

- The connection between [rr] and Italian restaurant names was not very robust, however.

Experiment 2

- Question: the association between [rr] and Italian names was not quite as robust as the other two effects.
- Control items (containing voiced obstruent geminates) were associated with Italian names with the probability that is much higher than chance.
- Maybe Japanese speakers do not want to use those names with voiced obstruent geminates for French restaurant names.

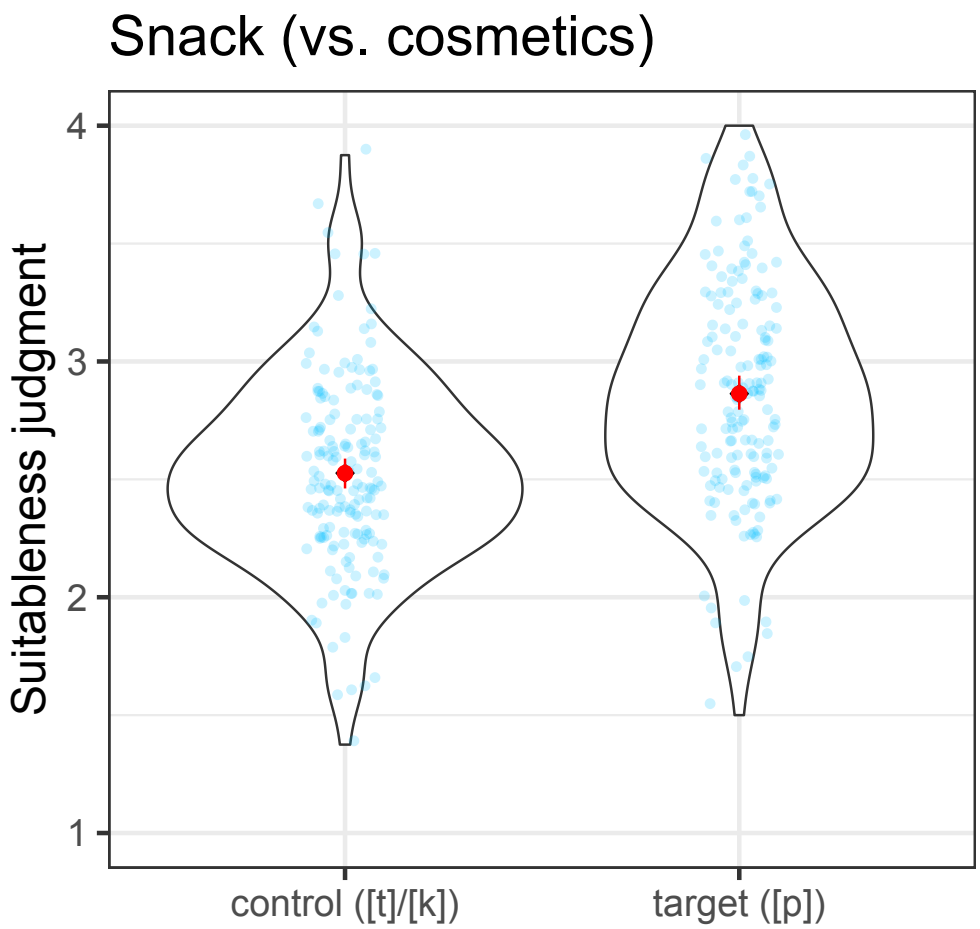


Experiment 2

- This time, we presented each stimulus and asked how suitable each name is for the three respective target categories on a 4-point scale.

- E.g. How suitable is [pariko] as a snack name? How suitable is [bohho] as a German person name?

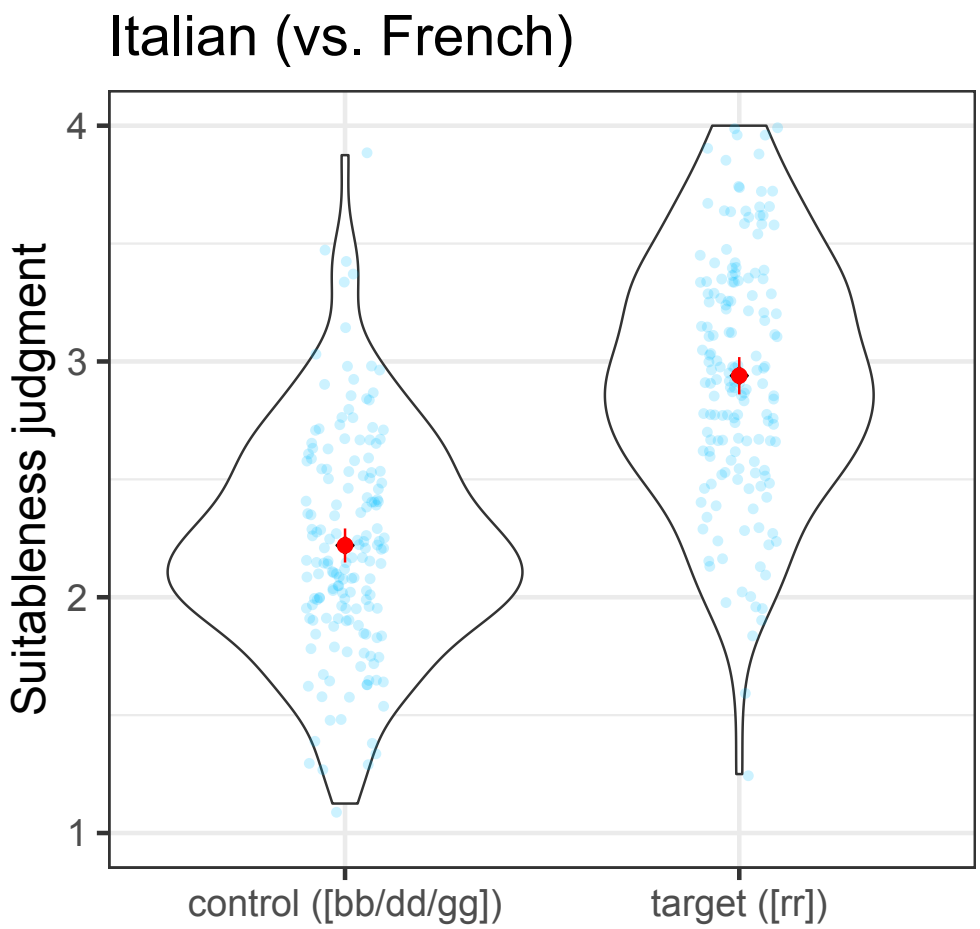
- 162 native speakers of Japanese.
- Experiment 1 contained [geppe] as one of the control items, but we excluded it because it may have reminded the Japanese speakers of [gebberusu], a (in)famous historical German figure.
- We used an ordinal logistic regression this time (Bayesian, mixed effects).



Results

- Those items that contain [p] were judged to be more suitable for snack names than the control items.

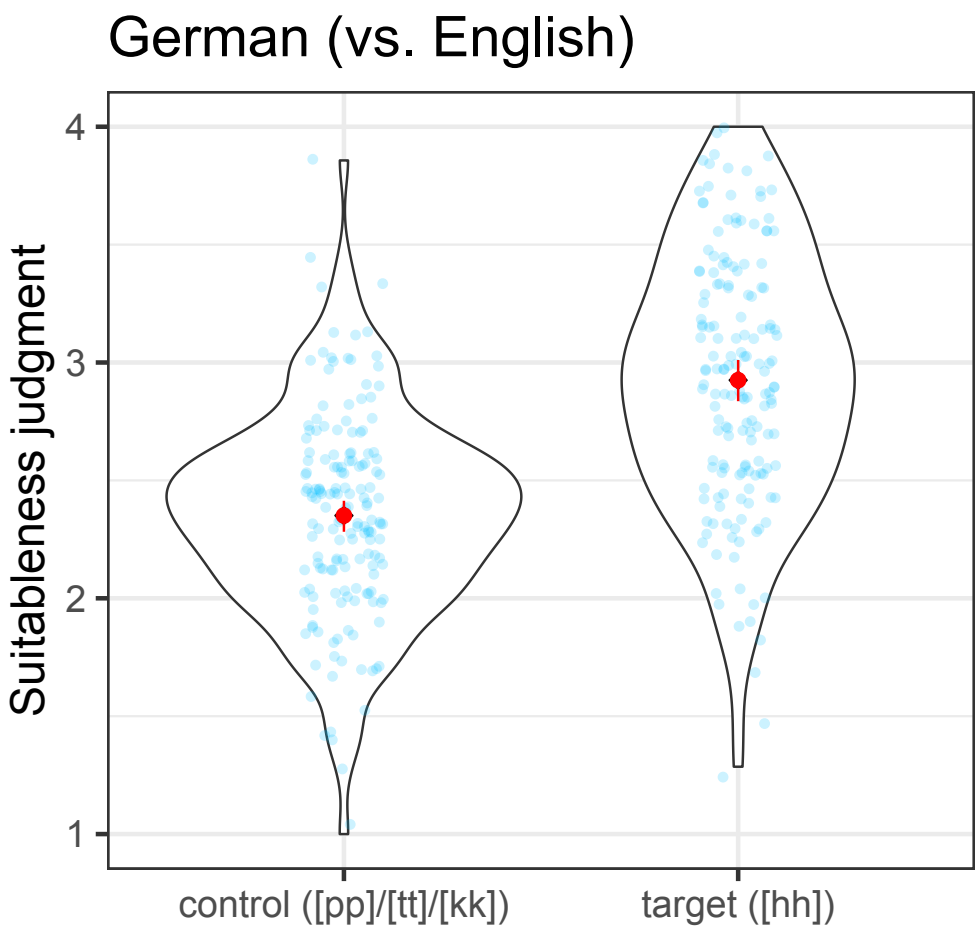
- $p(\beta 1>0) = 0.91$.



Results

- Those items that contain [rr] were judged to be more suitable for Italian names than the control items.

- $p(\beta 1>0) = 1$.



Results

- Those items that contain [hh] were more judged to be suitable for German names than the control items.

- $p(\beta 1>0) = 1$.

Conclusions

- Japanese speakers can infer, based a particular phonotactic cue, a very specific subpart of the lexicon, that is something as specific as “snack names” or “Italian names” or “German names”.

- In this sense, the traditional lexical stratification into native, SJ, loanwords and mimetics does not suffice.

- This is compatible with the view put forward by Gouskova et al. (2015) and others.

What does this mean theory-wise?

- Express(X) (Alderete & Kochetov 2017): Use a particular type of sound to express a particular type of meaning.

- Express(PAL <for small>): Use palatal consonants or [i] -- sounds with high F2 -- to express smallness.

- Express ([p] <for snack>), Express([rr] <for Italian>), Express([hh] <for German>).

/tariaterre/ (Italian)	Express([rr] <for Italian>)	*[rr]	Faith
=> [tariaterre]		*	
[tariatere]	*!		*

/tariaterre/	Express([rr] <for Italian>)	*[rr]	Faith
[tariaterre]		*!	
=>[tariatere]			*

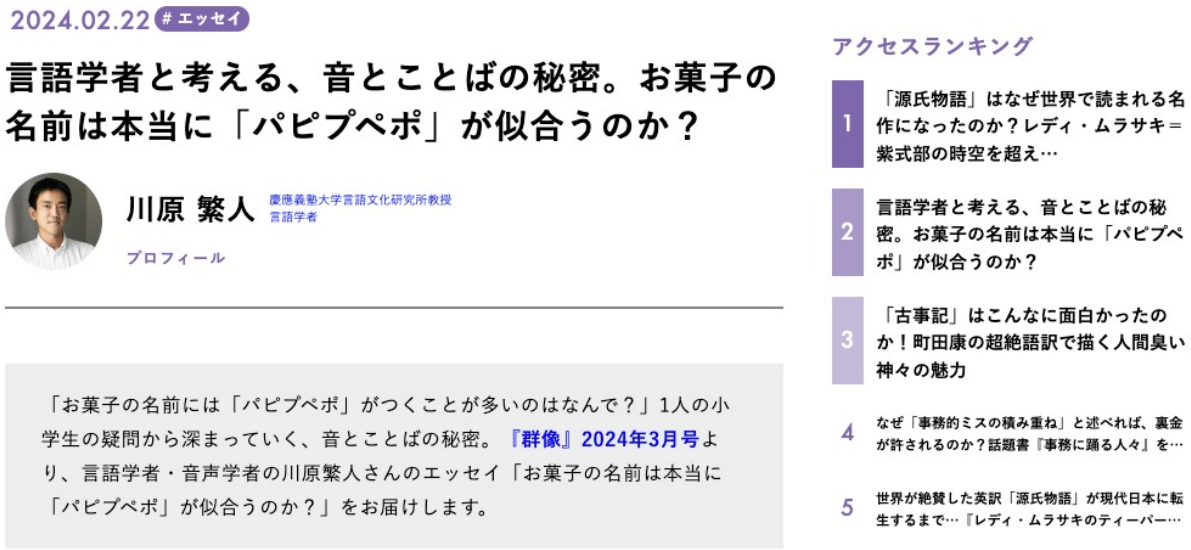
- Express(X) constraints are formulated as markedness constraints, reflecting the intuition that “we should use sound X to sound like Y”.
- But we can formulate the same constraint as a faithfulness constraint.

And…

- I was not wrong about the “[p]=snacks” association.

- With hindsight, however, I shouldn’t have allowed the publisher to use the title “Why X”, without examining the truth value of “X”.

- But at least I am being honest here. I admit that I made a mistake and examined my “hidden” assumption.
- It is amazing how often publishers use the “Why X”/“How-X” titles, presupposing that X is truth-conditionally correct. As a linguist, we should probably make a public warning about this practice.



2024.02.22 # エッセイ

言語学者と考える、音とことばの秘密。お菓子の名前は本当に「パピプペポ」が似合うのか？

川原 繁人 慶應義塾大学言語文化研究所教授 言語学者

プロフィール

「お菓子の名前には「パピプペポ」がつくことが多いのはなんで？」1人の小学生の疑問から深まっていく、音とことばの秘密。『群像』2024年3月号より、言語学者・音声学者の川原繁人さんのエッセイ「お菓子の名前は本当に「パピプペポ」が似合うのか？」をお届けします。

アクセスランキング

1 「源氏物語」はなぜ世界で読まれる名作になったのか？レディ・ムラサキ＝紫式部の時空を超え…

2 言語学者と考える、音とことばの秘密。お菓子の名前は本当に「パピプペポ」が似合うのか？

3 「古事記」はこんなに面白かったのか！町田康の超絶語訳で描く人間臭い神々の魅力

4 なぜ「事務的ミスの積み重ね」と述べれば、裏金が許されるのか？話題書『事務に踊る人々』を…

5 世界が絶賛した英訳「源氏物語」が現代日本に転生するまで…『レディ・ムラサキのティーパー…

Thank you