

Avoiding phonological markedness via word ordering in French and Italian

Katherine Blake

7 January 2022
LSA 96th Annual Meeting



Cornell University

Overview

I present evidence that shows that phonologically marked structures like vowel hiatus are avoided using word ordering.

Outline

- **Background:** languages and previous work
- **Methodology:** corpora, semantic clustering, statistical analysis
- **French:** phonology and results
- **Italian:** phonology and results
- **Discussion:** why some structures may be avoided but not others
- **Conclusion:** summary and key takeaways

Background

Variations in {noun, adjective} ordering

- Default ordering of modified noun phrases in French and Italian is POSTNOMINAL [noun adjective]
- PRENOMINAL [adjective noun] is also available for some adjectives
- Adjectives of interest: those that can occur in **both** positions

Type	French	Italian
(1) Strictly prenominal	?? <i>une maison belle</i> ~ <i>une belle maison</i> 'a nice house'	* <i>uno ritardo mero</i> ~ <i>un mero ritardo</i> 'a mere delay'
(2) Strictly postnominal	<i>l'industrie chimique</i> ~ * <i>la chimique industrie</i> 'chemical industry'	<i>un ingegnere elettronico</i> ~ * <i>un elettronico ingegnere</i> 'an electrical engineer'
(3) Flexible	<i>une maison magnifique</i> ~ <i>une magnifique maison</i> 'a beautiful house'	<i>il contributo prezioso</i> ~ <i>il prezioso contributo</i> 'precious contribution'

Previous related work

Binomials: Stress clash, word length, vowel quality, sonorancy affect binomial ordering in English (Bolinger, 1962; Pinker & Birdsong, 1979; Morgan, 2016; Ryan, 2019a)

Sentence formation: Stress clash, long consonant clusters, sibilant clash, geminates, vowel hiatus, bad sonority, and nasal-voiceless consonant clusters all affect syntactic structure of sentences in English (Breiss & Hayes, 2019)

Topicalization: NPs can be shifted for topicalization only if they are at least two phonological phrases (Serbo-Croatian; Zec & Inkelas, 1990)

Noun-adjective ordering: Nasal-nasal and nasal-voiceless consonant sequences avoided in {noun, adjective} ordering (Tagalog; Shih & Zuraw, 2017)

Research question

Do we see similar evidence for phonological markedness avoidance effects on {noun, adjective} ordering in French and Italian, with adjectives that are flexible?

- If so, are only those phonologically-marked phenomena that are avoided with phonological repairs also avoided with syntactic repairs?
(Shih & Zuraw, 2017)

Methodology

Corpora

- Common Voice (Mozilla) speech corpora

French	Italian
747 hours	288 hours
130,000 {noun, adjective} pairs	73,000 {noun, adjective} pairs

- Sentences tagged for POS using `spaCy`
- Phonological information provided by Lexique 3 (New et al., 2004) and PhonItalia (Goslin et al., 2014)

Semantic clustering

- Adjectives may have a difference in meaning depending on their position relative to the noun (Cinque, 2010)

Italian

a. un uomo povero
a man poor

'a poor man' (not rich)

b. un povero uomo
a poor man

'a pitiful man'

French

a. un homme grand
a man big

'a tall man'

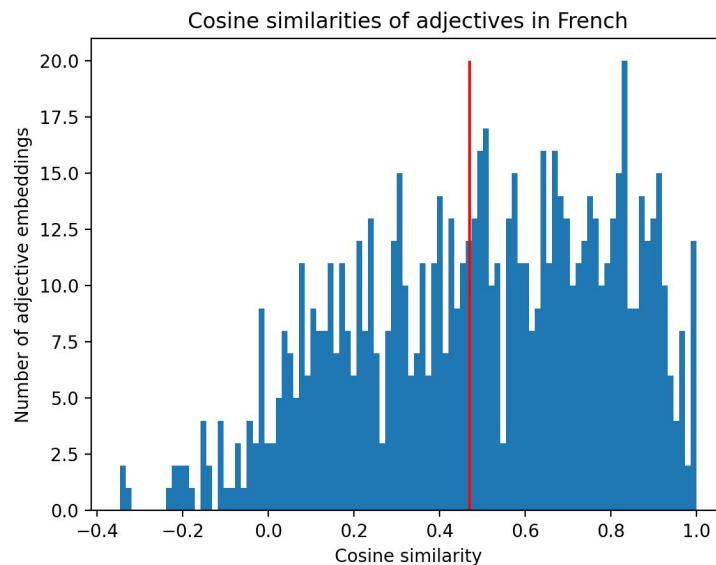
b. un grand homme
a big man

'a great man'

- Corpus data was split into two datasets to help control for this

Semantic clustering

- Separate prenominal and postnominal word embeddings were computed
 - Bag of words \rightarrow PPMI \rightarrow PCA (128 dimensions)
- Cosine similarity was measured between each prenominal and postnominal embedding of the same adjective (e.g., *grand* N and N *grand*)
- Gaussian mixture model with $k=2$ fit to the cosine similarities to determine a cutoff threshold for *similar* and *dissimilar* adjectives



Mixed-effects logistic regression models

- Predict the ordering of all {noun, adjective} pairs in the corpus that contain a flexible adjective (token-level)
- Fixed effects: phonological constraints, and relative frequency
 - Phonological constraints:
1=well-formed in prenominal, -1=well-formed in postnominal, 0=no preference
 - **Relative frequency**: a number between 0.0 and 0.5 that corresponds to how flexible a pair is
- Full random effects structure, with random slopes
- Separate models for *similar* and *dissimilar* adjectives

Hypotheses and predictions

- **HYPOTHESIS 1:** Only those phonologically-marked phenomena that are avoided with phonological repairs may also be avoided with syntactic repairs.
 - *Prediction:* Only those effects will be significantly positive in a model predicting ordering of {noun, adjective} pairs
- **HYPOTHESIS 2:** Phonological effects on ordering are stronger if semantic differences between orders are minimal.
 - *Prediction:* In the semantically-similar model, phonological effects will have a larger positive coefficient or be significant compared to the semantically-dissimilar model

Phonological constraints

Vowel hiatus vowel-vowel sequences across a syllable boundary

Stress clash: two adjacent prominent syllables

Stress lapse: three adjacent unstressed syllables

Clusters with mismatching voicing: voiced-voiceless or voiceless-voiced

Clusters with matching place of articulation: labial-labial, coronal-coronal, etc.

Long before short sequences: word 1 greater syllable count than word 2

French

Phonological constraints: French

Vowel hiatus

- Liaison repairs hiatus after certain words, under certain conditions (Tranel, 1995)
 - a. grands hommes
[gʁɑ̃ zɑ̃m]
'big men'
 - b. avions américains
[avjɑ̃ zamɛʁikɛ̃]
'American planes'

Phonological constraints: French

Clusters with mismatching voicing

- Regressive voicing assimilation occurs between obstruents of different voicing specifications (Snoeren & Segui, 2003)
 - a. une jupe droite
[yn ʒy^b dʁwat]
'a straight skirt'
 - b. une robe claire
[yn ʁo^p klɛʁ]
'a light dress'

Phonological constraints: French

Length

- Preference for shorter words before longer words in noun-adjective ordering (Forsgren, 1978; Thuilier, 2012)
 - a. un air avide
a air greedy
'a greedy air'
 - b. un avide hippopotame
a greedy hippopotamus
'a greedy hippopotamus'

Phonological constraints: French

OCP-Place

- Dispreference for obstruent sequences at the same place of articulation, but no strong evidence in Standard French

Stress constraints

- Not possible in French, no word-level stress

Phonological constraints: French

CONSTRAINT	ACTIVE STATUS
CLASH	Not possible.
LAPSE	Not possible.
HIATUS	Active across word boundaries (<i>liaison</i> ; (Tranel, 1995))
VOICE	Active across word boundaries (<i>regressive assimilation</i> ;; (Snoeren and Segui, 2003))
OCP-PLACE	Not active.
LENGTH	Active for noun-adjective pairs (Forsgren, 1978; Thuilier, 2012)

Results & Discussion: French

French

Dataset with semantically *dissimilar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	-6.16104	0.29052	-21.207	< 2e-16 ***
Constraint: HIATUS	0.97032	0.28915	3.356	0.000791 ***
Constraint: VOICE	0.21504	0.08444	2.547	0.010879 *
Constraint: OCP	0.17366	0.11287	1.539	0.123911
Constraint: LENGTH	0.01622	0.07945	0.204	0.838232
RELATIVE FREQUENCY	3.90285	0.23226	16.804	< 2e-16 ***

Dataset with semantically *similar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	0.7418	0.5281	1.405	0.1601
Constraint: HIATUS	1.5249	0.3807	4.005	6.19e-05 ***
Constraint: VOICE	0.2456	0.1468	1.672	0.0945 .
Constraint: OCP	0.1316	0.1688	0.780	0.4355
Constraint: LENGTH	0.2973	0.1165	2.551	0.0107 *
RELATIVE FREQUENCY	-2.6553	0.5080	-5.227	1.73e-07 ***

Predictions

- Significant positive coefficients for HIATUS, VOICE, and LENGTH
- No effect of OCP
- Stronger/present effects in the *similar* model

French

Dataset with semantically *dissimilar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	-6.16104	0.29052	-21.207	< 2e-16 ***
Constraint: HIATUS	0.97032	0.28915	3.356	0.000791 ***
Constraint: VOICE	0.21504	0.08444	2.547	0.010879 *
Constraint: OCP	0.17366	0.11287	1.539	0.123911
Constraint: LENGTH	0.01622	0.07945	0.204	0.838232
RELATIVE FREQUENCY	3.90285	0.23226	16.804	< 2e-16 ***

Dataset with semantically *similar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	0.7418	0.5281	1.405	0.1601
Constraint: HIATUS	1.5249	0.3807	4.005	6.19e-05 ***
Constraint: VOICE	0.2456	0.1468	1.672	0.0945 .
Constraint: OCP	0.1316	0.1688	0.780	0.4355
Constraint: LENGTH	0.2973	0.1165	2.551	0.0107 *
RELATIVE FREQUENCY	-2.6553	0.5080	-5.227	1.73e-07 ***

Hiatus

- Significantly positive coefficient in both models, larger in the *similar* model



French

Dataset with semantically *dissimilar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	-6.16104	0.29052	-21.207	< 2e-16 ***
Constraint: HIATUS	0.97032	0.28915	3.356	0.000791 ***
Constraint: VOICE	0.21504	0.08444	2.547	0.010879 *
Constraint: OCP	0.17366	0.11287	1.539	0.123911
Constraint: LENGTH	0.01622	0.07945	0.204	0.838232
RELATIVE FREQUENCY	3.90285	0.23226	16.804	< 2e-16 ***

Dataset with semantically *similar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	0.7418	0.5281	1.405	0.1601
Constraint: HIATUS	1.5249	0.3807	4.005	6.19e-05 ***
Constraint: VOICE	0.2456	0.1468	1.672	0.0945 .
Constraint: OCP	0.1316	0.1688	0.780	0.4355
Constraint: LENGTH	0.2973	0.1165	2.551	0.0107 *
RELATIVE FREQUENCY	-2.6553	0.5080	-5.227	1.73e-07 ***

Length

- Significantly positive coefficient only in the *similar* model



French

Dataset with semantically *dissimilar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	-6.16104	0.29052	-21.207	< 2e-16 ***
Constraint: HIATUS	0.97032	0.28915	3.356	0.000791 ***
Constraint: VOICE	0.21504	0.08444	2.547	0.010879 *
Constraint: OCP	0.17366	0.11287	1.539	0.123911
Constraint: LENGTH	0.01622	0.07945	0.204	0.838232
RELATIVE FREQUENCY	3.90285	0.23226	16.804	< 2e-16 ***

Dataset with semantically *similar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	0.7418	0.5281	1.405	0.1601
Constraint: HIATUS	1.5249	0.3807	4.005	6.19e-05 ***
Constraint: VOICE	0.2456	0.1468	1.672	0.0945 .
Constraint: OCP	0.1316	0.1688	0.780	0.4355
Constraint: LENGTH	0.2973	0.1165	2.551	0.0107 *
RELATIVE FREQUENCY	-2.6553	0.5080	-5.227	1.73e-07 ***

Voice

- Significantly positive only in the *dissimilar* model



French

Dataset with semantically *dissimilar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	-6.16104	0.29052	-21.207	< 2e-16 ***
Constraint: HIATUS	0.97032	0.28915	3.356	0.000791 ***
Constraint: VOICE	0.21504	0.08444	2.547	0.010879 *
Constraint: OCP	0.17366	0.11287	1.539	0.123911
Constraint: LENGTH	0.01622	0.07945	0.204	0.838232
RELATIVE FREQUENCY	3.90285	0.23226	16.804	< 2e-16 ***

Dataset with semantically *similar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	0.7418	0.5281	1.405	0.1601
Constraint: HIATUS	1.5249	0.3807	4.005	6.19e-05 ***
Constraint: VOICE	0.2456	0.1468	1.672	0.0945 .
Constraint: OCP	0.1316	0.1688	0.780	0.4355
Constraint: LENGTH	0.2973	0.1165	2.551	0.0107 *
RELATIVE FREQUENCY	-2.6553	0.5080	-5.227	1.73e-07 ***

OCP

- Insignificant in both models



French

Dataset with semantically *dissimilar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	-6.16104	0.29052	-21.207	< 2e-16 ***
Constraint: HIATUS	0.97032	0.28915	3.356	0.000791 ***
Constraint: VOICE	0.21504	0.08444	2.547	0.010879 *
Constraint: OCP	0.17366	0.11287	1.539	0.123911
Constraint: LENGTH	0.01622	0.07945	0.204	0.838232
RELATIVE FREQUENCY	3.90285	0.23226	16.804	< 2e-16 ***

Dataset with semantically *similar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	0.7418	0.5281	1.405	0.1601
Constraint: HIATUS	1.5249	0.3807	4.005	6.19e-05 ***
Constraint: VOICE	0.2456	0.1468	1.672	0.0945 .
Constraint: OCP	0.1316	0.1688	0.780	0.4355
Constraint: LENGTH	0.2973	0.1165	2.551	0.0107 *
RELATIVE FREQUENCY	-2.6553	0.5080	-5.227	1.73e-07 ***

Relative frequency

- Significant in both models: positive in *dissimilar* model and negative in *similar* model

French

As predicted

- HIATUS: avoiding hiatus is correlated with the order that avoids it in both models, more strongly in the *similar* model
- LENGTH: preference for short before long only significant in the *similar* model
- OCP: no effect

Not as predicted

- VOICE effect only in the *dissimilar* model

French

VOICE is able to be repaired without word-order manipulation,

whereas HIATUS and LENGTH are not

Regressive voicing assimilation

a. une jupe droite
[yn ʒy^b dʁwat]
'a straight skirt'

b. une robe claire
[yn ʁo^p klɛʁ]
'a light dress'

h aspiré words

a. grands haricot
[gʁɑ̃ Øaʁiko]
'big beans'

b. grandes haches
[gʁɑ̃d Øaʃ]
'big axes'

Epenthesis not possible

a. oiseau artificiel
[wazo aʁtifikjɛl]
'artificial bird'

Italian

Phonological constraints: Italian

Stress clash

- Clash may be repaired with stress retraction or initial-consonant doubling (Nespor & Vogel, 1979)
 - a. città vecchia
city old
 - b. città vvecchia
city old
'old city'

Phonological constraints: Italian

Stress lapse

- Lapse may be repaired with beat addition (Nespor & Vogel, 1989)

a. vecchia **alleanza**
old alliance
'old alliance'

Phonological constraints: Italian

Length

- Restrictions on sentence structure based on length of constituents (Cardinaletti, 2010)
 - a. Il partito di maggioranza fece poi la stessa proposta.
the party of majority made then the same proposal
 - b. **La stessa proposta** fece poi il partito di maggioranza.
the same proposal made then the party of majority
'The majority party then made the same proposal (not a similar one)'
 - c. ***La stessa proposta** fece poi Gianni/lui.

Phonological constraints: Italian

Vowel hiatus

- No strong evidence for phonological repairs of vowel-vowel sequences in Standard Italian

Voice and OCP

- Not possible in Italian, final consonants are extremely marginal

Phonological constraints: Italian

CONSTRAINT	ACTIVE STATUS
CLASH	Active across word boundaries (<i>retraction or doubling</i> ; (Nespor and Vogel, 1979))
LAPSE	Active across word boundaries (<i>beat addition</i> ; (Nespor and Vogel, 1989))
HIATUS	Not active.
VOICE	Not possible.
OCP-PLACE	Not possible.
LENGTH	Active for larger constituents (<i>object-verb-subject order</i> ; (Cardinaletti, 2010))

Results & Discussion: Italian

Italian

Dataset with semantically *dissimilar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	-2.10476	0.23881	-8.814	< 2e-16 ***
Constraint: CLASH	0.26362	0.31272	0.843	0.39924
Constraint: LAPSE	-0.29341	0.09038	-3.247	0.00117 **
Constraint: HIATUS	-0.89913	0.13404	-6.708	1.98e-11 ***
Constraint: LENGTH	0.39835	0.07318	5.444	5.22e-08 ***
RELATIVE FREQUENCY	1.56574	0.20247	7.733	1.05e-14 ***

Dataset with semantically *similar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	3.8106	1.1618	3.280	0.001039 **
Constraint: CLASH	0.1924	1.1880	0.162	0.871343
Constraint: LAPSE	0.2260	0.3298	0.685	0.493171
Constraint: HIATUS	-2.1790	0.5686	-3.832	0.000127 ***
Constraint: LENGTH	0.7740	0.2259	3.427	0.000611 ***
RELATIVE FREQUENCY	-6.7861	1.2392	-5.476	4.35e-08 ***

Predictions

- Significant positive coefficients for CLASH, LAPSE, and LENGTH
- No effect of HIATUS
- Stronger/present effects in the *similar* model

Italian

Dataset with semantically *dissimilar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	-2.10476	0.23881	-8.814	< 2e-16 ***
Constraint: CLASH	0.26362	0.31272	0.843	0.39924
Constraint: LAPSE	-0.29341	0.09038	-3.247	0.00117 **
Constraint: HIATUS	-0.89913	0.13404	-6.708	1.98e-11 ***
Constraint: LENGTH	0.39835	0.07318	5.444	5.22e-08 ***
RELATIVE FREQUENCY	1.56574	0.20247	7.733	1.05e-14 ***

Dataset with semantically *similar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	3.8106	1.1618	3.280	0.001039 **
Constraint: CLASH	0.1924	1.1880	0.162	0.871343
Constraint: LAPSE	0.2260	0.3298	0.685	0.493171
Constraint: HIATUS	-2.1790	0.5686	-3.832	0.000127 ***
Constraint: LENGTH	0.7740	0.2259	3.427	0.000611 ***
RELATIVE FREQUENCY	-6.7861	1.2392	-5.476	4.35e-08 ***

Length

- Significantly positive coefficient in both models, larger in the *similar* model



Italian

Dataset with semantically *dissimilar* adjectives

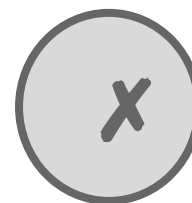
	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	-2.10476	0.23881	-8.814	< 2e-16 ***
Constraint: CLASH	0.26362	0.31272	0.843	0.39924
Constraint: LAPSE	-0.29341	0.09038	-3.247	0.00117 **
Constraint: HIATUS	-0.89913	0.13404	-6.708	1.98e-11 ***
Constraint: LENGTH	0.39835	0.07318	5.444	5.22e-08 ***
RELATIVE FREQUENCY	1.56574	0.20247	7.733	1.05e-14 ***

Dataset with semantically *similar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	3.8106	1.1618	3.280	0.001039 **
Constraint: CLASH	0.1924	1.1880	0.162	0.871343
Constraint: LAPSE	0.2260	0.3298	0.685	0.493171
Constraint: HIATUS	-2.1790	0.5686	-3.832	0.000127 ***
Constraint: LENGTH	0.7740	0.2259	3.427	0.000611 ***
RELATIVE FREQUENCY	-6.7861	1.2392	-5.476	4.35e-08 ***

Clash

- Insignificant in both models



Italian

Dataset with semantically *dissimilar* adjectives

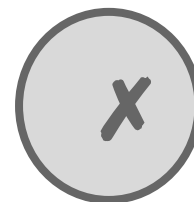
	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	-2.10476	0.23881	-8.814	< 2e-16 ***
Constraint: CLASH	0.26362	0.31272	0.843	0.39924
Constraint: LAPSE	-0.29341	0.09038	-3.247	0.00117 **
Constraint: HIATUS	-0.89913	0.13404	-6.708	1.98e-11 ***
Constraint: LENGTH	0.39835	0.07318	5.444	5.22e-08 ***
RELATIVE FREQUENCY	1.56574	0.20247	7.733	1.05e-14 ***

Dataset with semantically *similar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	3.8106	1.1618	3.280	0.001039 **
Constraint: CLASH	0.1924	1.1880	0.162	0.871343
Constraint: LAPSE	0.2260	0.3298	0.685	0.493171
Constraint: HIATUS	-2.1790	0.5686	-3.832	0.000127 ***
Constraint: LENGTH	0.7740	0.2259	3.427	0.000611 ***
RELATIVE FREQUENCY	-6.7861	1.2392	-5.476	4.35e-08 ***

Lapse

- Significantly negative in the *dissimilar* model, insignificant in *similar*



Italian

Dataset with semantically *dissimilar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	-2.10476	0.23881	-8.814	< 2e-16 ***
Constraint: CLASH	0.26362	0.31272	0.843	0.39924
Constraint: LAPSE	-0.29341	0.09038	-3.247	0.00117 **
Constraint: HIATUS	-0.89913	0.13404	-6.708	1.98e-11 ***
Constraint: LENGTH	0.39835	0.07318	5.444	5.22e-08 ***
RELATIVE FREQUENCY	1.56574	0.20247	7.733	1.05e-14 ***

Dataset with semantically *similar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	3.8106	1.1618	3.280	0.001039 **
Constraint: CLASH	0.1924	1.1880	0.162	0.871343
Constraint: LAPSE	0.2260	0.3298	0.685	0.493171
Constraint: HIATUS	-2.1790	0.5686	-3.832	0.000127 ***
Constraint: LENGTH	0.7740	0.2259	3.427	0.000611 ***
RELATIVE FREQUENCY	-6.7861	1.2392	-5.476	4.35e-08 ***

Hiatus

- Significantly negative in both models, larger effect in *similar* model



Italian

Dataset with semantically *dissimilar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	-2.10476	0.23881	-8.814	< 2e-16 ***
Constraint: CLASH	0.26362	0.31272	0.843	0.39924
Constraint: LAPSE	-0.29341	0.09038	-3.247	0.00117 **
Constraint: HIATUS	-0.89913	0.13404	-6.708	1.98e-11 ***
Constraint: LENGTH	0.39835	0.07318	5.444	5.22e-08 ***
RELATIVE FREQUENCY	1.56574	0.20247	7.733	1.05e-14 ***

Dataset with semantically *similar* adjectives

	ESTIMATE	STD. ERROR	Z VALUE	P VALUE
Intercept	3.8106	1.1618	3.280	0.001039 **
Constraint: CLASH	0.1924	1.1880	0.162	0.871343
Constraint: LAPSE	0.2260	0.3298	0.685	0.493171
Constraint: HIATUS	-2.1790	0.5686	-3.832	0.000127 ***
Constraint: LENGTH	0.7740	0.2259	3.427	0.000611 ***
RELATIVE FREQUENCY	-6.7861	1.2392	-5.476	4.35e-08 ***

Relative frequency

- Significant in both models: positive in *dissimilar* model and negative in *similar* model

Italian

As predicted

- LENGTH: preference for short before long is significant in both models, stronger in the *similar* model

Not as predicted

- CLASH and LAPSE are not avoided with word-order manipulation
- HIATUS is likelier to be tolerated than avoided

Italian

CLASH and LAPSE are also able to be repaired without word-order manipulation,
whereas LENGTH is not

Clash repairs

- a. città vecchia
city old
- b. città vvecchia
city old
'old city'

Lapse repair

- a. vecchia alleanza
old alliance
'old alliance'

Italian

HIATUS may be the result of the phonological shape of words in Italian

- 73% of {noun, adjective} pairs that can violate hiatus do so
- 71% of these violations are POSTNOMINAL [noun adjective] order, meaning there is a vowel-final noun before a vowel-initial adjective
- 98% of nouns in the corpus are vowel-final, and 18% of adjectives are vowel-initial

Constraint includes penalization of high vowel-vowel sequences, which may surface as glide-vowel and not violate hiatus

Discussion

Inactive phonological constraints

- VOICE, CLASH, and LAPSE were all predicted to have a significantly positive effect on word-ordering, but did not
- For all three constraints, a phonological repair in the output order was available (compared to HIATUS and LENGTH)
- Word-order manipulation is not a preferred repair strategy to phonology

Relative frequency

Included as a control effect for degree of how fixed a pair is in one order

Positive in *dissimilar* models

- Flexible pairs are likelier to be prenominal, which is the position to which specific or special meanings of an adjective are usually attributed

Negative in *similar* models

- Flexible noun-adjective pairs are likelier to be postnominal, which is the default order

Takeaways

Summary & key findings

Do we see evidence for phonological markedness avoidance effects on {noun, adjective} ordering in French and Italian, with adjectives that are flexible?

→ Yes, with two stipulations

(1) phonological constraints must be otherwise ACTIVE in the language

(2) phonological repair strategies may outcompete word-order manipulation

Summary & key findings

Are phonological effects stronger when semantic difference is weaker?

→ Yes

Relevant phonological effects were present only in, or had a larger coefficient in, regression models with pairs that included adjectives with a smaller semantic difference between their prenominal and postnominal positions

Thank you

Dissertation committee

- Drs. Marten van Schijndel, Helena Aparicio, Draga Zec and Abigail Cohn

Cornell lab groups

- Computational Psycholinguistics Discussions (C.Psyd)
- Cornell Phonetics Lab (PLab)



Citations

Bolinger, D. L. (1962). Binomials and pitch accent. *Lingua*, 11:34–44.

Breiss, C. and Hayes, B. (2019). Phonological markedness effects in sentence formation. Ms., University of California, Los Angeles.

Cardinaletti, A. (2010). On a (wh-) moved topic in Italian, compared to Germanic. *Advances in comparative Germanic syntax*, pages 3–40.

Cinque, G. (2010). *The syntax of adjectives: A comparative study*, volume 57. MIT press.

Forsgren, M. (1978). *La place de l'adjectif épithète en français contemporain: étude quantitative et sémantique*. PhD thesis, Acta Universitatis Upsaliensis.

Goslin, J., Galluzzi, C., and Romani, C. (2014). Phonitalia: a phonological lexicon for Italian. *Behavior research methods*, 46(3):872–886.

Knittel, M. L. (2005). Some remarks on adjective placement in the French NP. *Probus*, 17(2):185–226.

Martin, A. (2011). Grammars leak: Modeling how phonotactic generalizations interact within the grammar. *Language*, 87(4):751–770.

Morgan, E. I. P. (2016). *Generative and item-specific knowledge of language*. University of California, San Diego.

Nespor, M. and Vogel, I. (1979). Clash avoidance in Italian. *Linguistic Inquiry*, 10(3):467–482.

Citations

Nespor, M. and Vogel, I. (1989). On clashes and lapses. *Phonology*, 6(1):69–116.

New, B., Pallier, C., Brysbaert, M., and Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, 36(3):516–524.

Pinker, S. and Birdsong, D. (1979). Speakers' sensitivity to rules of frozen word order. *Journal of Verbal Learning and Verbal Behavior*, 18(4):497–508.

Ryan, K. M. (2019a). Prosodic end-weight reflects phrasal stress. *Natural Language & Linguistic Theory*, 37(1):315–356.

Shih, S. S. and Zuraw, K. (2017). Phonological conditions on variable adjective and noun word order in Tagalog. *Language*, 93(4):e317–e352.

Snoeren, N. D. and Segui, J. (2003). A voice for the voiceless: Voice assimilation in French. In *Proceedings of the 15th International Congress of Phonetic Sciences*, pages 2325–2328. Citeseer.

Thuilier, J. (2012). *Contraintes préférentielles et ordre des mots en français*. PhD thesis, Université Paris-Diderot-Paris VII.

Tranel, B. (1995). Current issues in French phonology: Liaison and position theories. *The Handbook of Phonological Theory*, pages 798–816.

Zec, D. and Inkelas, S. (1990). Prosodically constrained syntax. *The phonology-syntax connection*, pages 365–378.