



Université  
de Paris



# Quantifying near-homophony induced by French liaison

Victor Antoine

Laboratoire de Sciences Cognitives et Psycholinguistique (LSCP)

Supervised by:

Rory Turnbull (Newcastle University)

Sharon Peperkamp (PSL University)



DEC

DÉPARTEMENT  
D'ÉTUDES  
COGNITIVES



## Research question

French liaison : pronunciation of an otherwise silent underlying word-final consonant ([g], [n], [t], [b], [p], [z]) when the following word is vowel-initial





## Research question

French liaison : pronunciation of an otherwise silent underlying word-final consonant ([g], [n], [t], [ʁ], [p], [z]) when the following word is vowel-initial



Word 1

dernier<sup>r</sup>

Word 2

chat

[pronunciation]

[d<sup>ʁ</sup>nje.<sup>ʃ</sup>a]





# Research question

French liaison : pronunciation of an otherwise silent underlying word-final consonant ([g], [n], [t], [ʁ], [p], [z]) when the following word is vowel-initial



Word 1

dernier<sup>r</sup>

dernier<sup>r</sup>



Word 2

chat

<sup>a</sup>chat

[pronunciation]

[dɛʁ<sup>n</sup>.nje.<sup>r</sup>ʃa]

[dɛʁ<sup>n</sup>.nje.<sup>a</sup>ʃa]



# Research question

French liaison : pronunciation of an otherwise silent underlying word-final consonant ([g], [n], [t], [ʁ], [p], [z]) when the following word is vowel-initial

Word 1

dernier<sup>r</sup>

dernier<sup>r</sup>

dernier<sup>r</sup>

Word 2

chat

a<sup>r</sup>chat

rachat

[pronunciation]

[dɛʁ<sup>n</sup>.nje.<sup>r</sup>ʃa]

[dɛʁ<sup>n</sup>.nje.<sup>a</sup>ʃa]

[dɛʁ<sup>n</sup>.nje.<sup>r</sup>ʃa]





# Research question

French liaison : pronunciation of an otherwise silent underlying word-final consonant ([g], [n], [t], [ʁ], [p], [z]) when the following word is vowel-initial



Word 1

dernier<sup>r</sup>

dernier<sup>r</sup>

dernier<sup>r</sup>

Word 2

chat

<sup>a</sup>chat

rachat

[pronunciation]

[dɛʁ.nje.ʃa]

[dɛʁ.nje.ʁa.ʃa]

[dɛʁ.nje.ʁa.ʃa]

- Liaison-induced near-homophonous sequences (Spinelli et al., 2002; Spinelli et al., 2003)



# Research question

French liaison : pronunciation of an otherwise silent underlying word-final consonant ([g], [n], [t], [ʁ], [p], [z]) when the following word is vowel-initial



Word 1

dernier<sup>r</sup>

dernier<sup>r</sup>

dernier<sup>r</sup>

Word 2

chat

<sup>a</sup>chat

rachat

[pronunciation]

[dɛʁ.nje.ʃa]

[dɛʁ.nje.ʁa.ʃa]

[dɛʁ.nje.ʁa.ʃa]

- Liaison-induced near-homophonous sequences (Spinelli et al., 2002; Spinelli et al., 2003)
- Introduction of lexical segmentation difficulties (Spinelli et al., 2003)



# Research question

French liaison : pronunciation of an otherwise silent underlying word-final consonant ([g], [n], [t], [ʁ], [p], [z]) when the following word is vowel-initial



Word 1

dernier<sup>r</sup>

dernier<sup>r</sup>

dernier<sup>r</sup>

Word 2

chat

<sup>a</sup>chat

rachat

[pronunciation]

[dɛʁ.nje.ʃa]

[dɛʁ.nje.ʁa.ʃa]

[dɛʁ.nje.ʁa.ʃa]

- Liaison-induced near-homophonous sequences (Spinelli et al., 2002; Spinelli et al., 2003)
- Introduction of lexical segmentation difficulties (Spinelli et al., 2003)
- Avoidance of challenges generated by homophones found in several languages (Dautriche et al., 2018)





## Research question

French liaison : pronunciation of an otherwise silent underlying word-final consonant ([g], [n], [t], [ʁ], [p], [z]) when the following word is vowel-initial

Word 1	Word 2	[pronunciation]
dernier <sup>r</sup>	chat	[dɛʁ.nje.ʃa]
dernier <sup>r</sup>	a <sup>a</sup> chat	[dɛʁ.nje.ʁa.ʃa]
dernier <sup>r</sup>	rachat	[dɛʁ.nje.ʁa.ʃa]

- Liaison-induced near-homophonous sequences (Spinelli et al., 2002; Spinelli et al., 2003)
- Introduction of lexical segmentation difficulties (Spinelli et al., 2003)
- Avoidance of challenges generated by homophones found in several languages (Dautriche et al., 2018)

**Is French structured in such a way as to reduce liaison-induced near-homophony ?**





# Number of confusing doublets

Using the electronic dictionary Lexique:

1. Words\_1

e.g. *dernier*





# Number of confusing doublets

Pre-registered analysis

Using the electronic dictionary Lexique:

1. Words\_1
2. Minimal pairs of Words\_2

e.g. *dernier*

e.g. *achat/rachat*





# Number of confusing doublets

Pre-registered analysis

Using the electronic dictionary Lexique:

1. Words\_1
2. Minimal pairs of Words\_2
3. Confusing doublets

e.g. *dernier*

e.g. *achat/rachat*

e.g. {*dernier achat / dernier rachat*}





# Number of confusing doublets

Pre-registered analysis

Using the electronic dictionary Lexique:

1. Words\_1
2. Minimal pairs of Words\_2
3. Confusing doublets
4. Removed ungrammatical ones

e.g. *dernier*

e.g. *achat/rachat*

e.g. {*dernier achat* / *dernier rachat*}

e.g. {\**dernier achètent* / \**dernier rachètent*}





# Number of confusing doublets

Pre-registered analysis

Using the electronic dictionary Lexique:

1. Words\_1 e.g. *dernier*
2. Minimal pairs of Words\_2 e.g. *achat/rachat*
3. Confusing doublets e.g. {*dernier achat / dernier rachat*}
4. Removed ungrammatical ones e.g. {*\*dernier achètent / \*dernier rachètent*}



**Comparison: alternative versions of French with different liaison consonants**





# Number of confusing doublets

Pre-registered analysis

Using the electronic dictionary Lexique:

1. Words\_1 e.g. *dernier*
2. Minimal pairs of Words\_2 e.g. *achat/rachat*
3. Confusing doublets e.g. {*dernier achat / dernier rachat*}
4. Removed ungrammatical ones e.g. {*\*dernier achètent / \*dernier rachètent*}



## Comparison: alternative versions of French with different liaison consonants

- Same Words\_1 but substituted liaison consonant e.g. *dernier* [b] → [f]





# Number of confusing doublets

Pre-registered analysis

Using the electronic dictionary Lexique:

1. Words\_1 e.g. *dernier*
2. Minimal pairs of Words\_2 e.g. *achat/rachat*
3. Confusing doublets e.g. {*dernier achat / dernier rachat*}
4. Removed ungrammatical ones e.g. {*\*dernier achètent / \*dernier rachètent*}



## Comparison: alternative versions of French with different liaison consonants

- Same Words\_1 but substituted liaison consonant e.g. *dernier* [b] → [f]
- Same sorting to remove ungrammatical ones e.g. *dernier* + masc. sing. noun







# Number of confusing doublets

Pre-registered analysis

Using the electronic dictionary Lexique:

1. Words\_1
2. Minimal pairs of Words\_2
3. Confusing doublets
4. Removed ungrammatical ones

e.g. *dernier*

e.g. *achat/rachat*

e.g. {*dernier achat / dernier rachat*}

e.g. {*\*dernier achètent / \*dernier rachètent*}

## Comparison: alternative versions of French with different liaison consonants

- Same Words\_1 but substituted liaison consonant
- Same sorting to remove ungrammatical ones
- New confusing doublets

e.g. *dernier* [b] → [f]

e.g. *dernier* + masc. sing. noun

e.g. {*dernier acteur / dernier facteur*}





# Number of confusing doublets

Pre-registered analysis

Using the electronic dictionary Lexique:

1. Words\_1
2. Minimal pairs of Words\_2
3. Confusing doublets
4. Removed ungrammatical ones

e.g. *dernier*

e.g. *achat/rachat*

e.g. {*dernier achat* / *dernier rachat*}

e.g. {*\*dernier achètent* / *\*dernier rachètent*}



## Comparison: alternative versions of French with different liaison consonants

- Same Words\_1 but substituted liaison consonant
- Same sorting to remove ungrammatical ones
- New confusing doublets

e.g. *dernier* [b] → [f]

e.g. *dernier* + masc. sing. noun

e.g. {*dernier acteur* / *dernier facteur*}



[b]

Confusing doublets 'real' French  
[b] {*dernier achat* / *dernier rachat*}

Number 'real' French  
[b] x



# Number of confusing doublets

Pre-registered analysis

Using the electronic dictionary Lexique:

1. Words\_1
2. Minimal pairs of Words\_2
3. Confusing doublets
4. Removed ungrammatical ones

e.g. *dernier*

e.g. *achat/rachat*

e.g. {*dernier achat / dernier rachat*}

e.g. {*\*dernier achètent / \*dernier rachètent*}

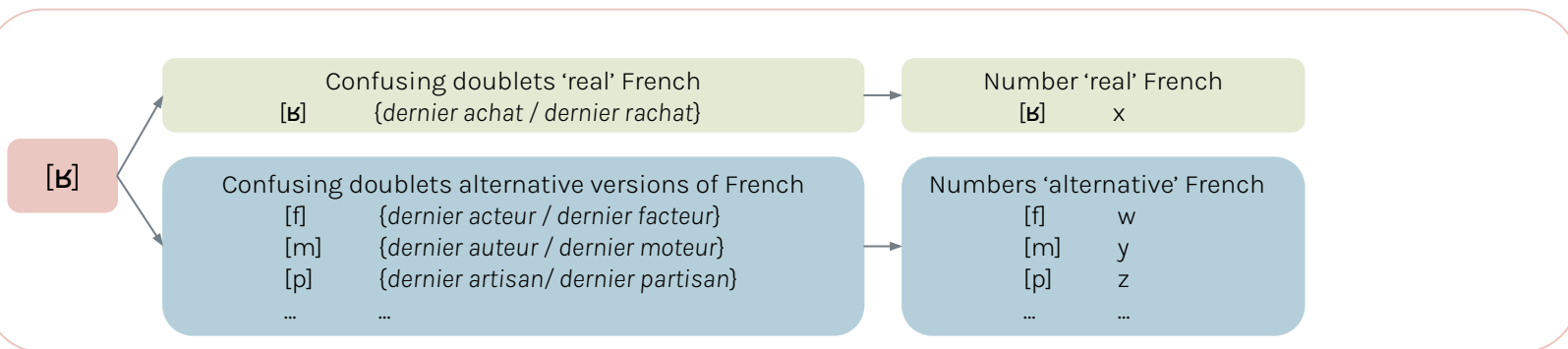
## Comparison: alternative versions of French with different liaison consonants

- Same Words\_1 but substituted liaison consonant
- Same sorting to remove ungrammatical ones
- New confusing doublets

e.g. *dernier* [ʁ] → [f]

e.g. *dernier* + masc. sing. noun

e.g. {*dernier acteur / dernier facteur*}





# Number of confusing doublets

Pre-registered analysis

Using the electronic dictionary Lexique:

1. Words\_1
2. Minimal pairs of Words\_2
3. Confusing doublets
4. Removed ungrammatical ones

e.g. *dernier*

e.g. *achat/rachat*

e.g. {*dernier achat / dernier rachat*}

e.g. {*\*dernier achètent / \*dernier rachètent*}



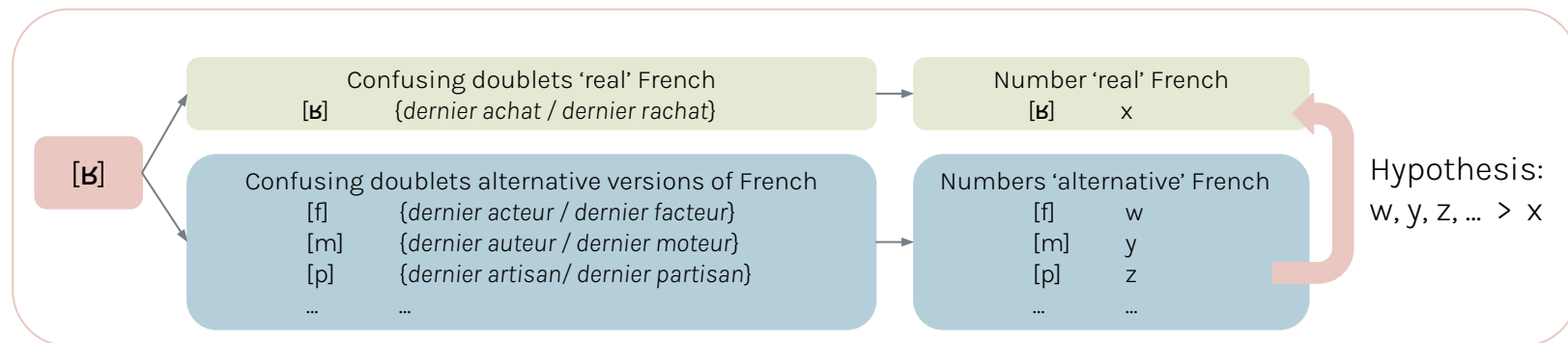
## Comparison: alternative versions of French with different liaison consonants

- Same Words\_1 but substituted liaison consonant
- Same sorting to remove ungrammatical ones
- New confusing doublets

e.g. *dernier* [b] → [f]

e.g. *dernier* + masc. sing. noun

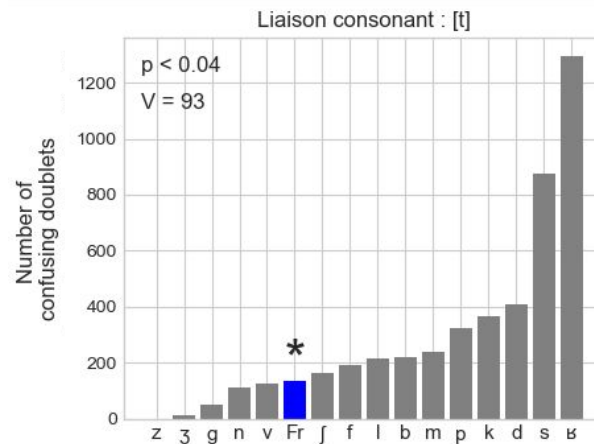
e.g. {*dernier acteur / dernier facteur*}





# Number of confusing doublets

Pre-registered analysis



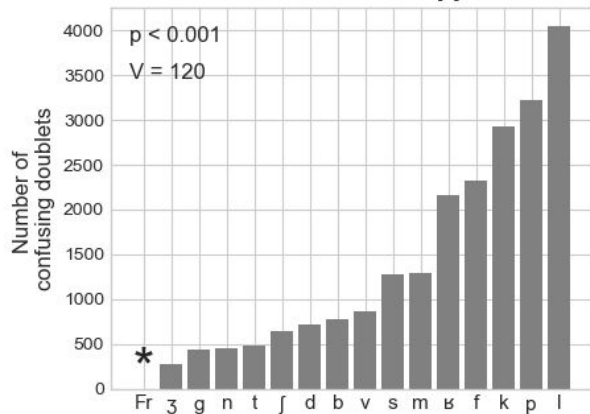


# Number of confusing doublets

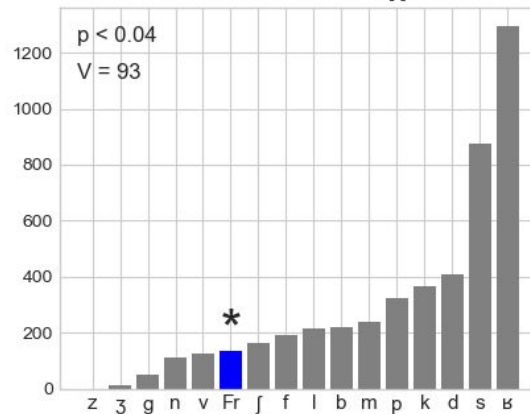
Pre-registered analysis



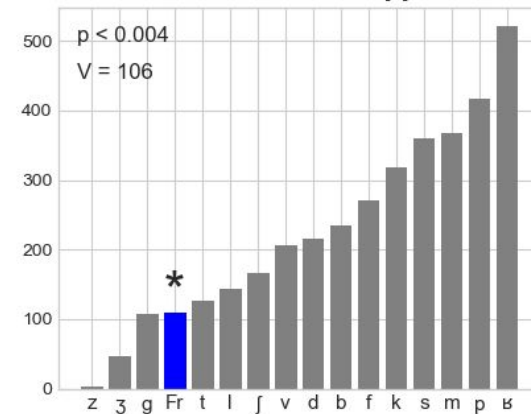
Liaison consonant : [z]



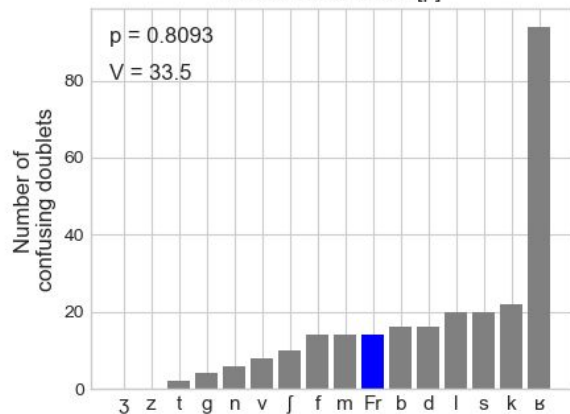
Liaison consonant : [t]



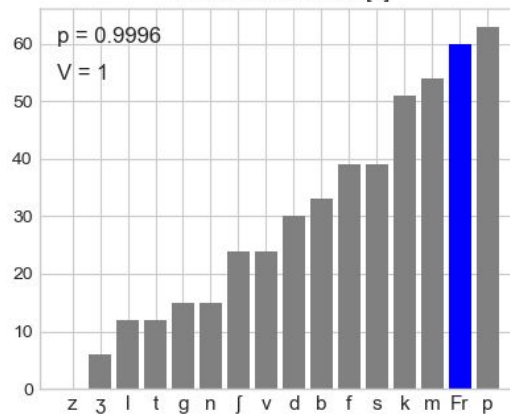
Liaison consonant : [n]



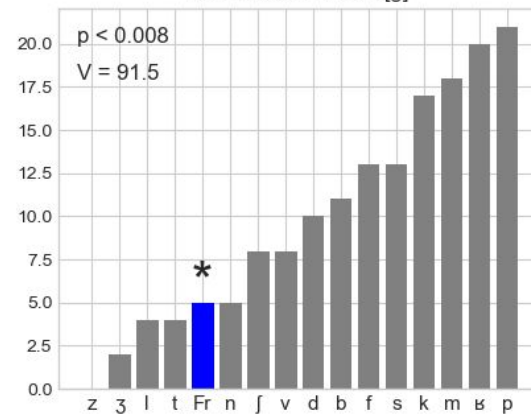
Liaison consonant : [p]



Liaison consonant : [ʁ]



Liaison consonant : [g]





# Number of confusing doublets



- Significant results for [n], [t], [z] and [g]
- Non-significant results for [b], [p]
- Median N confusing doublets 'alternative' French > N 'real' French for **4/6 liaison cons.**
- Including the most prevalent liaison consonants



Liaison consonant	[g]	[p]	[b]	[n]	[t]	[z]
Number of Words_1	1	2	3	17	1419	3217



# Number of confusing doublets

Pre-registered analysis



- Significant results for [n], [t], [z] and [g]
- Non-significant results for [b], [p]
- Median N confusing doublets 'alternative' French > N 'real' French for **4/6 liaison cons.**
- Including the most prevalent liaison consonants



Liaison consonant	[g]	[p]	[b]	[n]	[t]	[z]
Number of Words_1	1	2	3	17	1419	3217



Other relevant parameters to consider

- Semantic analysis

*Do the doublets designate close concepts?*

*Would other liaison cons. be better?*

e.g. {petit ami / petit amis}

e.g. petit ami ≠ petit amis

- Frequency analysis

*Are the members' frequency values close to each other?*

*Would other liaison cons. be better?*

e.g. petit ami > petit amis







# Number of 'troublesome' doublets

Exploratory analysis

For all doublets (real or alternative French), retrieval of frequency data using Google Ngram Viewer.





# Number of 'troublesome' doublets

Exploratory analysis

For all doublets (real or alternative French), retrieval of frequency data using Google Ngram Viewer.

Assumed doublet is troublesome if:

Parameter:

Frequency data  
troublesome doublet

1. at least one member is highly  
frequent

e.g {est en / est tant}

## Doublet Sum

$$\text{Sum} = \text{freq}(a) + \text{freq}(b)$$

$$\text{Sum} = \text{freq}(\text{est en}) + \text{freq}(\text{est tant})$$

High Sum





# Number of 'troublesome' doublets

Exploratory analysis

For all doublets (real or alternative French), retrieval of frequency data using Google Ngram Viewer.

Assumed doublet is troublesome if:

Parameter:

Frequency data  
troublesome doublet

1. at least one member is highly frequent

## Doublet Sum

$$\text{Sum} = \text{freq}(a) + \text{freq}(b)$$

High Sum

e.g {est en / est tant}

$$\text{Sum} = \text{freq}(\text{est en}) + \text{freq}(\text{est tant})$$

2. both members are in the same frequency range

$$\text{Ratio} = \frac{\min(\text{freq}(a), \text{freq}(b))}{\max(\text{freq}(a), \text{freq}(b))}$$

Ratio close to 1

e.g [b] → [f]

{dernier acteur / dernier facteur}

$$\text{Ratio} = \frac{\text{freq}(\text{dernier acteur})}{\text{freq}(\text{dernier facteur})}$$



# Number of 'troublesome' doublets

**Statistical decision: troublesome doublet if its sum and ratio are in the top 50%**

$SUM \in \text{top 50\% sum}$     AND     $RATIO \in \text{top 50\% ratio}$





# Number of 'troublesome' doublets

Statistical decision: troublesome doublet if its sum and ratio are in the top 50%

$SUM \in \text{top 50\% sum}$  AND  $RATIO \in \text{top 50\% ratio}$



[b]

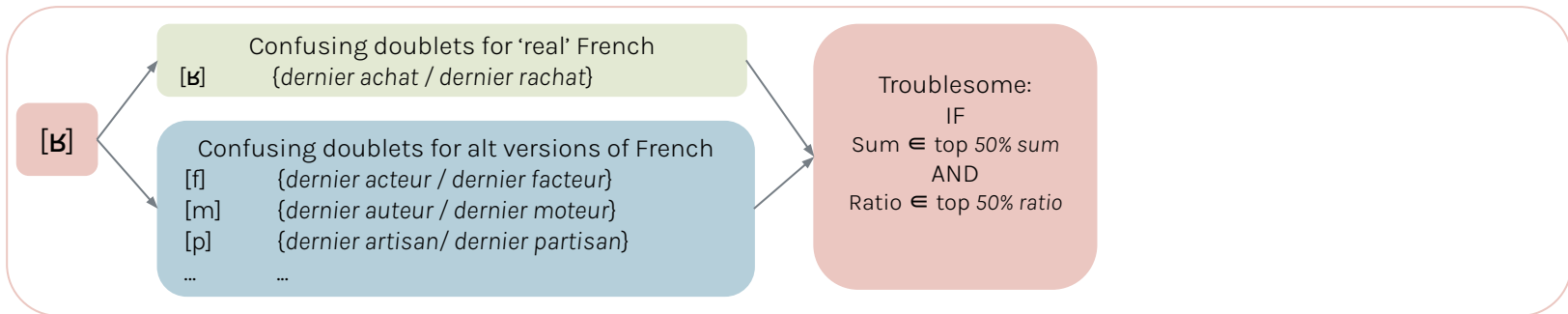
Confusing doublets for 'real' French  
[b] {dernier achat / dernier rachat}

Confusing doublets for alt versions of French  
[f] {dernier acteur / dernier facteur}  
[m] {dernier auteur / dernier moteur}  
[p] {dernier artisan / dernier partisan}  
... ..

# Number of 'troublesome' doublets

Statistical decision: troublesome doublet if its sum and ratio are in the top 50%

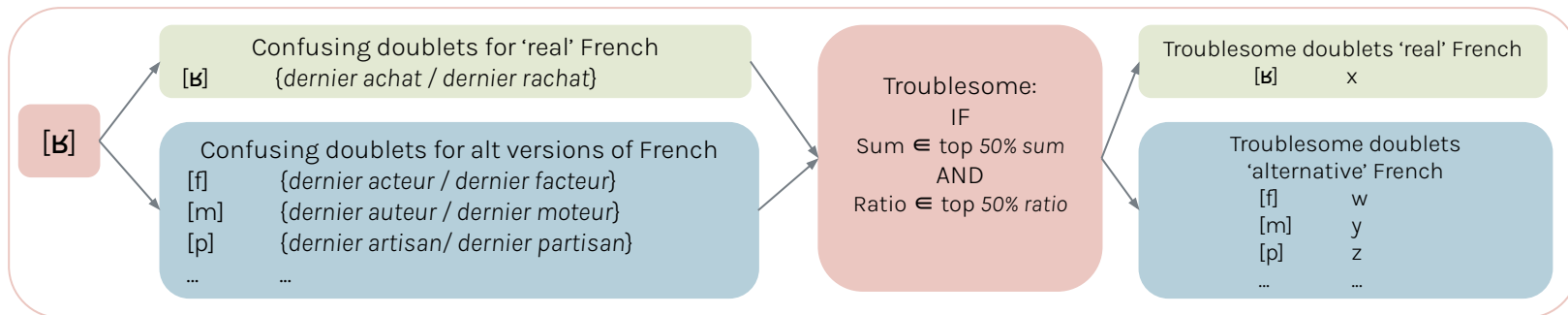
$SUM \in \text{top 50\% sum}$  AND  $RATIO \in \text{top 50\% ratio}$



# Number of 'troublesome' doublets

Statistical decision: troublesome doublet if its sum and ratio are in the top 50%

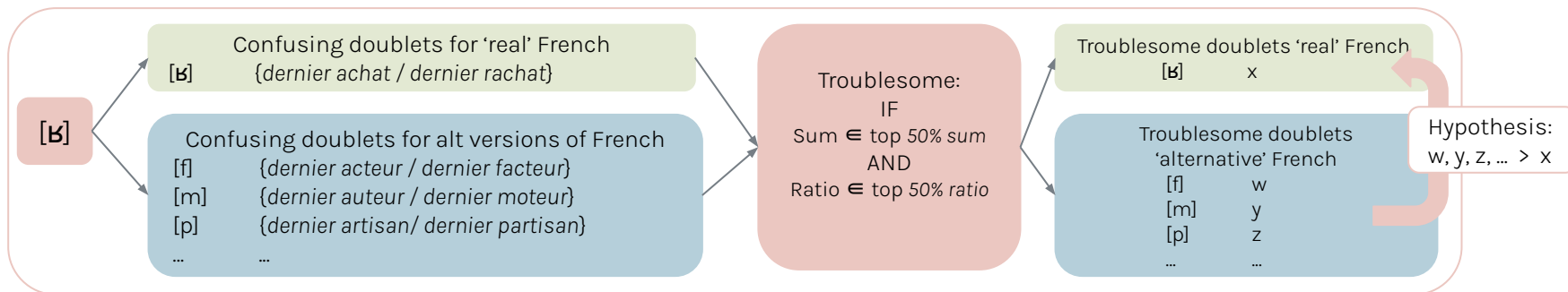
$SUM \in \text{top 50\% sum}$  AND  $RATIO \in \text{top 50\% ratio}$



# Number of 'troublesome' doublets

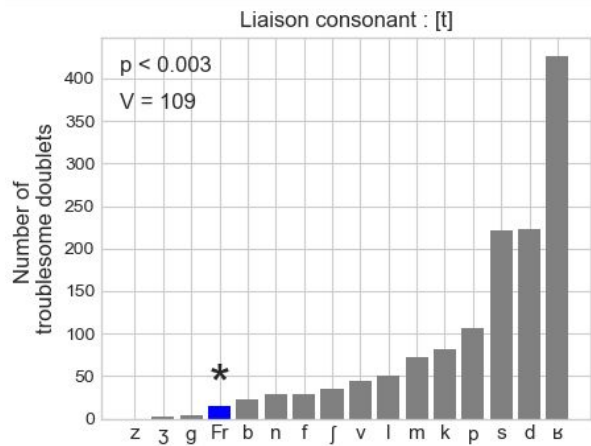
Statistical decision: troublesome doublet if its sum and ratio are in the top 50%

$$\text{SUM} \in \text{top 50\% sum} \quad \text{AND} \quad \text{RATIO} \in \text{top 50\% ratio}$$





# Number of 'troublesome' doublets

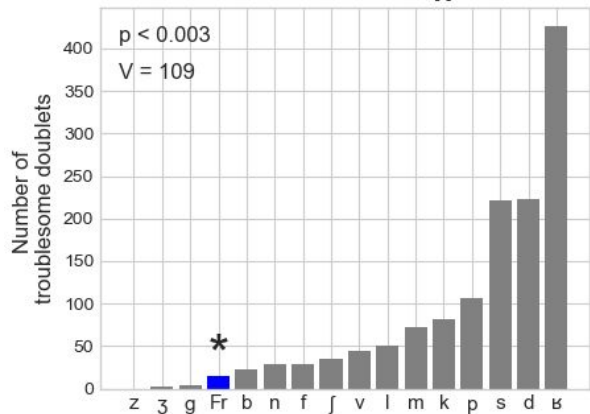




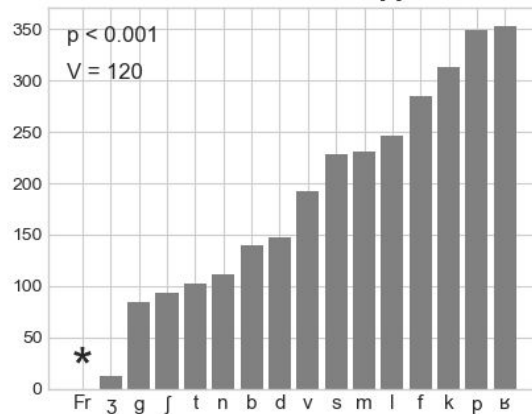
# Number of 'troublesome' doublets

Exploratory analysis

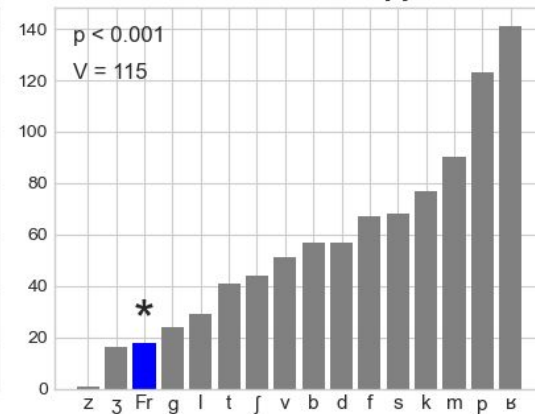
Liaison consonant : [t]



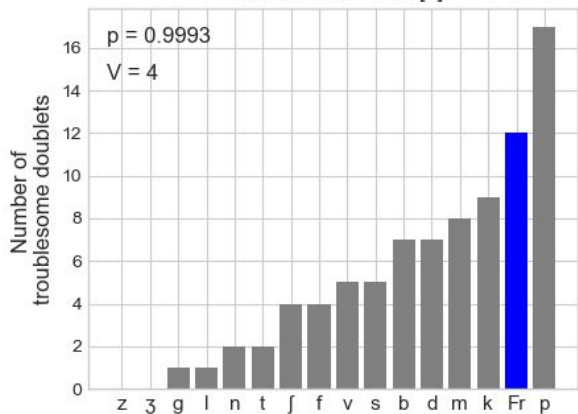
Liaison consonant : [z]



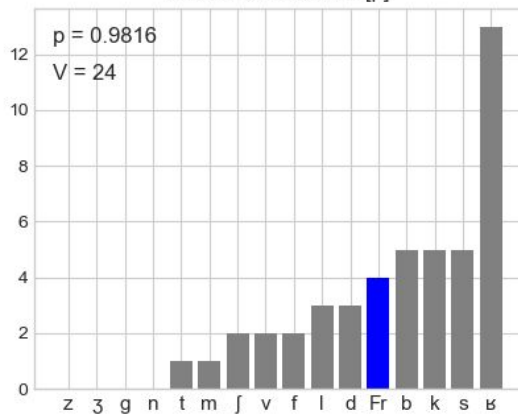
Liaison consonant : [n]



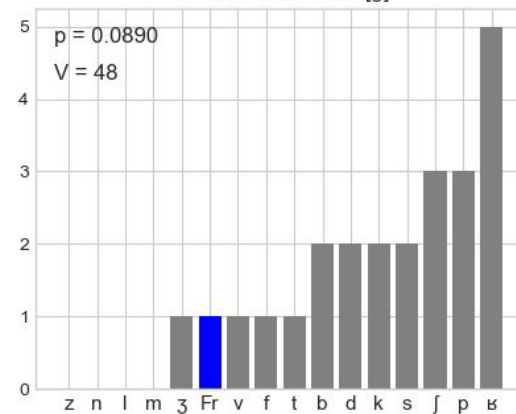
Liaison consonant : [ʁ]



Liaison consonant : [p]



Liaison consonant : [g]





## Discussion



- Significant results for [n], [t] and [z] in both analyses
- French - lexicon and grammar - structured to reduce liaison-induced difficulties





## Discussion



- Significant results for [n], [t] and [z] in both analyses
- French - lexicon and grammar - structured to reduce liaison-induced difficulties



- Subjective process
- Fine details not captured by current constraints  
(*premier habit* vs. *dernier habit* vs. *~léger habit*)





# Discussion



- Significant results for [n], [t] and [z] in both analyses
- French - lexicon and grammar - structured to reduce liaison-induced difficulties



- Subjective process
- Fine details not captured by current constraints  
(*premier habit* vs. *dernier habit* vs. *~léger habit*)



- Creation of a composite score
  - ✓ ○ Number of confusing doublets (raw number of near-homophonous sequences)
  - ✓ ○ Number of troublesome doublets (frequency analysis)
  - NEW ○ Constraints on what follows the doublets e.g. {*est au* / *est tôt*}
  - NEW ○ Semantic analysis e.g. *petit ami* ≠ *petit amis*
- Investigate experimentally 'real-life' difficulties of such doublets





# Discussion



- Significant results for [n], [t] and [z] in both analyses
- French - lexicon and grammar - structured to reduce liaison-induced difficulties



- Subjective process
- Fine details not captured by current constraints  
(*premier habit* vs. *dernier habit* vs. *~léger habit*)



- Creation of a composite score
  - ✓ ○ Number of confusing doublets (raw number of near-homophonous sequences)
  - ✓ ○ Number of troublesome doublets (frequency analysis)
  - NEW ○ Constraints on what follows the doublets e.g. {*est au* / *est tôt*}
  - NEW ○ Semantic analysis e.g. *petit ami* ≠ *petit tamis*
- Investigate experimentally 'real-life' difficulties of such doublets

