Networks and nondeterminism

Maria Copot Andrea D. Sims

The Ohio State University

Inflectional realisation

- Mapping a lexeme to the inflectional form that realises a paradigmatic cell.
  - F(lexeme, cell) $\rightarrow$ inflected form
  - F(BAKE, PRS.3SG) $\rightarrow$ *bakes*
- Nondeterminism of inflectional realisation arises when the function above does not yield a clear, unique output.

Conceptualising inflectional nondeterminism

- The process underlying inflectional realisation is about the interaction of an input with the wider system it is part of
- Nondeterminism is predicted to arise where the lexicon enters high-entropy configurations.

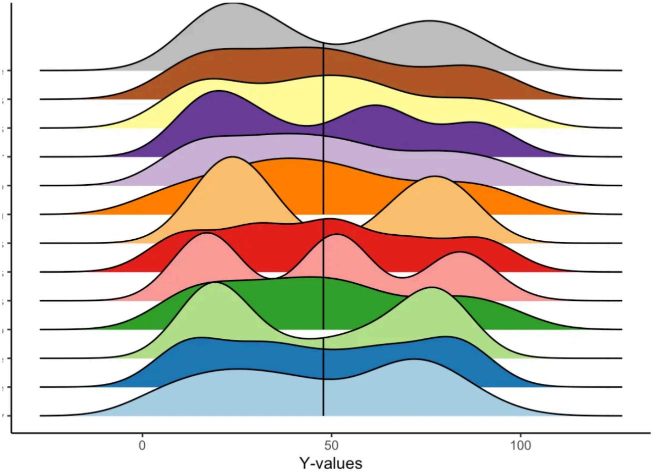
Information theory and nondeterminism

- **Information-theoretic** approaches to the lexicon are able to quantify the **uncertainty** involved in the realisation of a lexeme in a cell

Lexeme	PRS.3PL	PRS.2PL	π_1	π_2	*Freq*	*Surprisal*	*Entropy*
BAVER	bav	bave	✓		1	0	0
MORDRE	mɔʁd	mɔʁde	✓		1	0	0
PEINER	pɛn	pɛne	✓	✓	13	1.585	0.918
LEVER	lɛv	ləve	✓	✓	23	0.585	0.918
MENER	mɛn	məne	✓	✓	23	0.585	0.918

Entropy

- Entropy is an **average** of a transformed distribution of probabilities.
- Like all averages, it **masks** important things about the **underlying distribution**



Entropy

- Imagine a lexeme L that could take four patterns in cell C with different likelihoods:

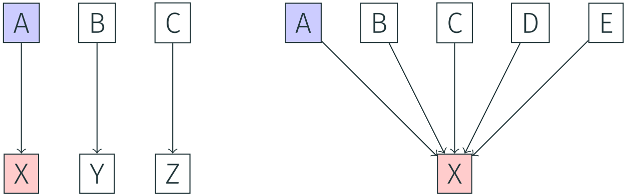
A	B	C	D	Entropy
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	1.75
$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	1.8

- Despite the differences in distribution…
  - The entropy values are very close
  - Disregard for which patterns are more similar

The importance of the underlying configuration in morphology

- Entropy of 0 (no uncertainty): a **strong implicative relationship** between two exponents, or a **default**?

H(“I am” $\rightarrow$ “he is”) H(“I swim” $\rightarrow$ “he swims”)



- Comparable **ambiguity** exist for cases of higher entropy.

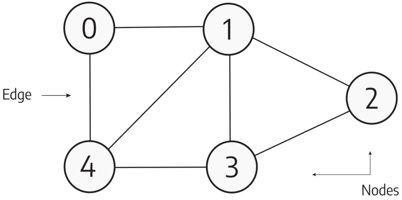
The underlying configurations of nondeterminism

- We present a configuration-forward approach to the lexicon, rooted in network theory
- We focus on two network measures - quantitative approach to outlining configurations predicted to yield inflectional nondeterminism
- Allows us to:
  - Make predictions about where speakers will behave nondeterministically and to what degree (work in progress!)
  - Query why nondeterminism occurs at a given locus
- Proof of concept of larger usefulness of networks to the study of the lexicon

Network science

A **network** is a mathematical structure characterised by

- nodes (objects)
- edges (connections)



An approach that centers **relationships between objects**.

Network science

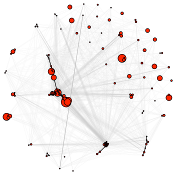
- Networks have been used to model system dynamics in the physical, biological and social sciences
- Limited adoption in modeling of lexical relations
  - e.g. Pham & Baayen (2015), Brown & Hippisley (2012)[1], Beniamine, (2016)[2], Sims (2020).

[1] Technically a tree
[2] Technically a lattice

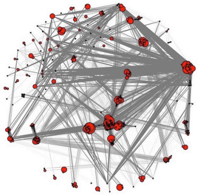
The linguistic systems - French verbs and BCMS nouns

An International Space Station view of the systems

French



BCMS



- A central cluster of microclasses.
- Most other classes are variations on the central cluster, characterised by the addition of unique exponents.

- No clearly defined central cluster, high interconnectedness
- Plenty of mixing and matching of exponents

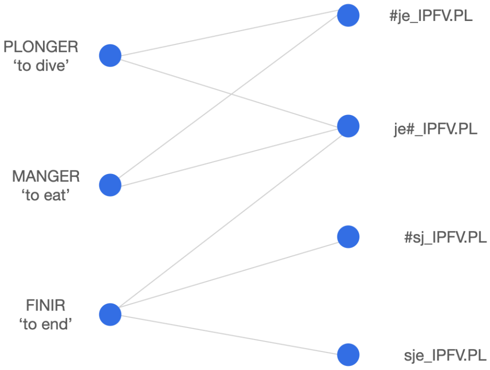
Setting up a network

Transposing an inflected lexicon to a network

1. Start with phonemically transcribed inflected lexicon
2. `morphalign` (Beniamine & Guzmán-Naranjo, 2020) to align inflected forms.
3. `setmorph` (Carroll & Beniamine, submitted; Beniamine & Carroll, 2023) to segment inflected forms into smallest discriminative units (lexeme-internal comparisons)
4. Convert discriminative units into exponents:
   - combined exponents that were reliably adjacent
   - combine exponents for which a combined variant has been identified elsewhere in the system.
5. To capture subexponent regularities, the exponent string is segmented into triphones
6. The triphones are marked for the paradigmatic cell they occur in

Bipartite network

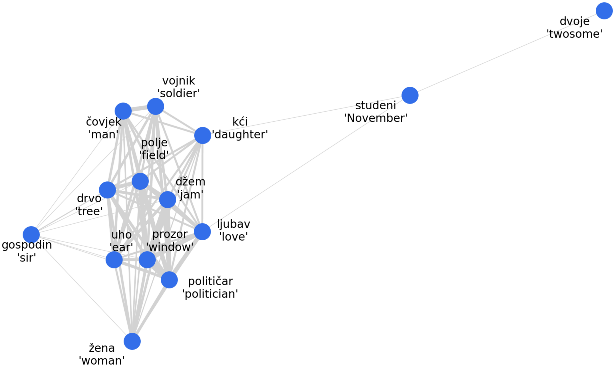
Two types of nodes: lexemes and exponent triphones.



	IPFV.2PL
PLONGER	plɔ̃ʒ-je
MANGER	mɑ̃ʒ-je
FINIR	fini-sje

One-mode network

- Project the bipartite network to a **one-mode network** (only one type of node: lexemes).
- Two lexemes are connected by an edge if they share **at least one exponent triphone**.
- Each edge is **weighted** by how many exponent triphones they share.



The graphs

We create **bipartite and one-mode** networks for the **French verbal system** (Vlexique 2.0, Beniamine et al. 2023) and the **BCMS nominal system** (UniMorph, Batsuren et al. 2022).

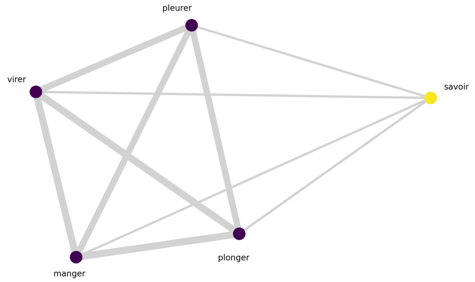
	French verbs	BCMS nouns
paradigm cells	52	12
unique lexemes	5.274	10.927
unique triphone + cell combinations	3.917	1.290
edges (bipartite)	605.896	202.999
edges (one mode)	13.865.187	56.057.017

Nondeterminism in networks

- We can expect nondeterminism to show up in a network as
  - A node being pulled in multiple directions, multiple possible analogical patterns
    - e.g. overabundance (Thornton, 2011)
  - A node being peripheral to the network, no good analogical pattern
    - e.g. lexemes peripheral to the system, susceptible to regularisation
- We can quantify these properties for each node thanks to the measures of BETWEENNESS CENTRALITY and CLOSENESS CENTRALITY

Closeness centrality - odd exponents

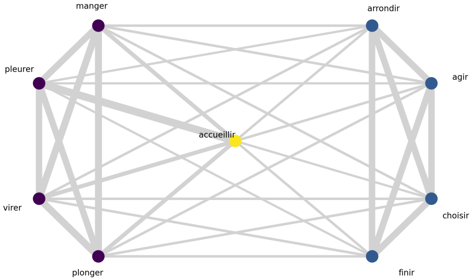
- Closeness (Bavelas, 1950) is a function of the identity of exponents
  - Are a lexeme's exponents well integrated with the rest of the system?
  - How unique are a lexeme's exponents?



- The average shortest distance between node N and any other node
- Nodes low in closeness are not well connected to the rest of the system and are peripheral to it

Betweenness centrality - oddness of exponent pairing

- Betweenness (Gross & Yellen, 2006) is a function of the **arrangement of exponents** of a lexeme
  - Is the lexeme pairing up exponents in unusual ways?



- The **number of shortest paths** between any two nodes in the network that **pass through node N**.
- Nodes high in betweenness are **bridges** for navigating the system

Lexemes with low closeness centrality

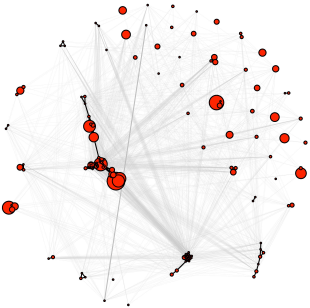
- French:
  - defective lexemes (e.g. ACCROIRE ‘make sb believe’, QUÉRIR ‘seek’)
  - suppletive verbs (e.g. AVOIR ‘to have’, ÊTRE ‘to be’, SAVOIR ‘to know’)
  - verbs with very rare or unique exponents (e.g. FAIRE ‘to do’)
- BCMS:
  - pluralia tanta neuters (e.g. PLUĆA ‘lungs’)
  - nouns with unusual exponent(s) (e.g. OBA ‘both’, GEN.PL *obaju*)
  - collectives (e.g. DUGMAD ‘buttons’)

Lexemes with high betweenness centrality

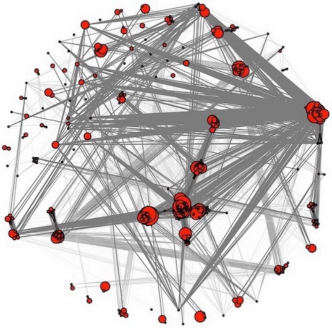
- French:
  - Verbs that mix exponents from the first and second conjugation (e.g. ACCUEILLIR 'welcome')
  - Third conjugation verbs with rare exponents that still take default exponents for part of their paradigm, making them a bridge between the third conjugation and the other two (e.g. VALOIR 'be worth', BOIRE 'drink')
- BCMS:
  - Large classes central to the system that mix exponents

Different types of items - why?

French

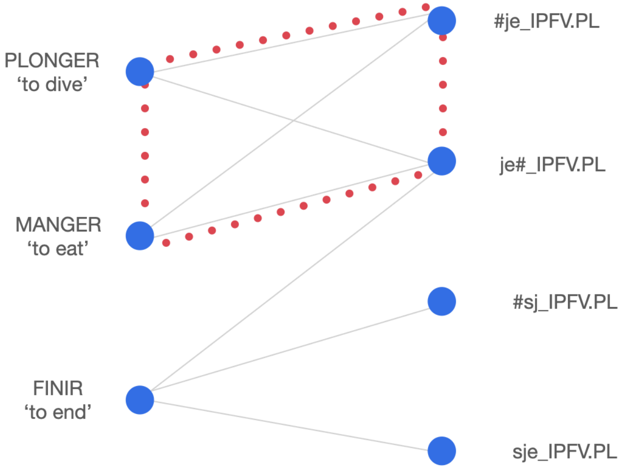


BCMS

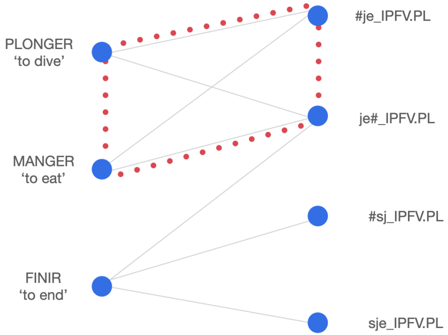


Square clustering

”The probability that my friends have common friends except me” - strength of joint probability of exponents.



Square clustering

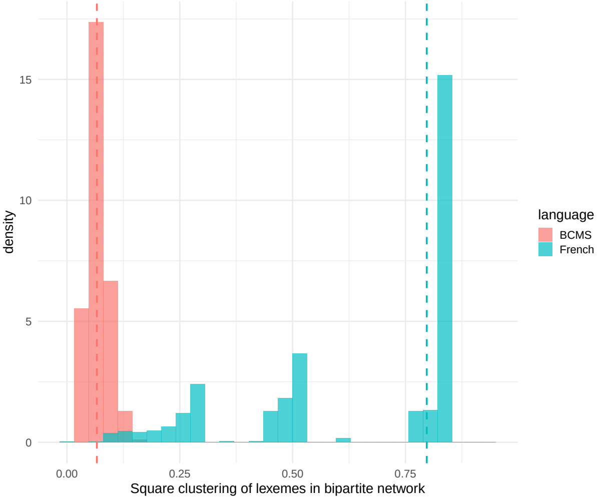


The network has one square. The SC coefficient measures how many squares a node is part of, out of the total possible number. In this network:

PLONGER	1/2 = 0.5
MANGER	1/2 = 0.5
FINIR	0

Square clustering in French verbs and BCMS nouns

The two systems have very **different configurations**: in French, exponents have much higher **joint probability** than in BCMS.



High betweenness centrality in BCMS nouns

- Lexemes high in betweenness centrality in BCMS come from large, frequent classes
    - Conceived of as core to the system.
- Because of low joint exponent probability overall, it is the lexemes most prototypical of the system that act as bridges between its different parts (unlike in French)
- Such lexemes have not been the focus of work on nondeterminism - perhaps because of their high type frequency?
    - Prediction to test: locus of nondeterminism.

Conclusion

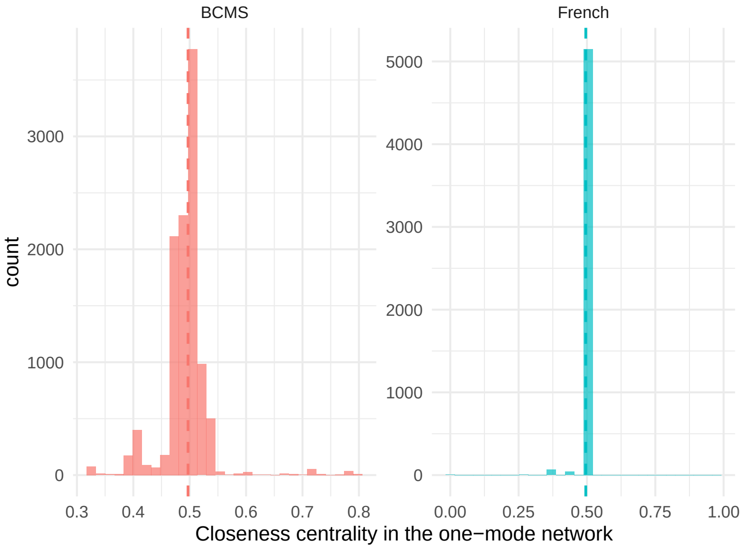
- A network-based framing of the lexicon gives us the tools to think in relational terms about both **local and global dynamics**
- The framework comes with an extensive **quantitative toolkit** for studying systems, and the role of individual items within them.
- Network theory allows us to investigate the different configurations that lead to **nondeterminism** in the context of their **broader system**
- Exploratory step in **broader research programme**: many other possible applications of network theory to the lexicon, come talk to us!

A networking opportunity

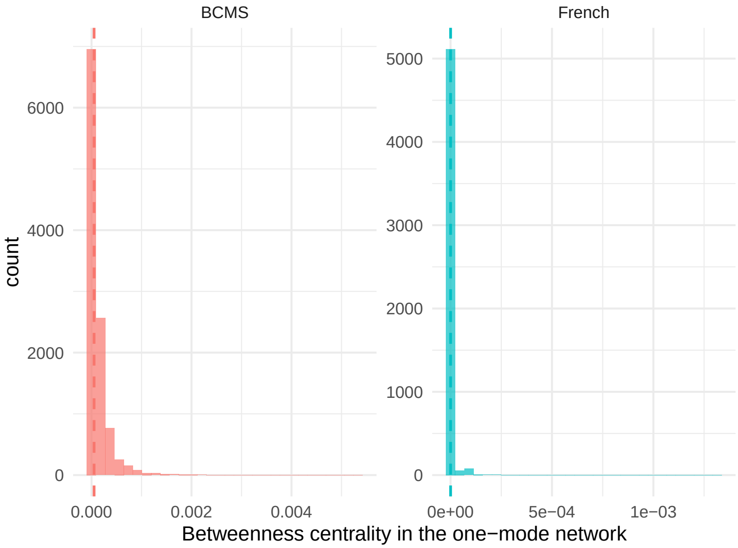
If you are interested in exploring the graphs or the lexeme/exponent measures you can find them on OSF:

Appendix

Closeness distribution



Betweenness distribution



Form predictability as average surprisal i

- We want to derive a quantitative measure of how much confidence a speaker may have that they are producing the appropriate form for a given paradigm cell, given knowledge of the rest of the paradigm.
- This is clearly a variant of the Paradigm Cell Filling Problem Ackerman09,Ackerman13.

- We rely on a purely word-based approach to the PCFP of Bonami16, using Beniamine18's (Beniamine18) Qumin package for all computations.

Form predictability as average surprisal ii

For each pair of cells (c, c') in the paradigm:

1. Assign each pair to an alternation pattern, optimizing alignments between pairs of words.

Form predictability as average surprisal iii

Lexeme	PRS.3PL	PRS.2PL
CROIRE	kʁwa	kwaje
BAVER	bav	bave
LEVER	lɛv	ləve
MENER	mɛn	məne
PEINER	pɛn	pɛne
MORDRE	mɔʁd	mɔʁde

⇒

Lexeme	PRS.3PL	PRS.2PL	Alternation
BAVER	bav	bave	π_1 : _ ⇌ _e / X^+ C_#
PEINER	pɛn	pɛne	
MORDRE	mɔʁd	mɔʁde	
LEVER	lɛv	ləve	π_2 : _ _ ⇌ _ə_e / X^+ _C_#
MENER	mɛn	məne	
CROIRE	kʁwa	kwaje	π_3 : _ ⇌ _je / X^+ wa_#

Form predictability as average surprisal iv

Lexeme	PRS.3PL	PRS.2PL	Alternation
BAVER	bav	bave	
PEINER	pɛn	pɛne	$\pi_1 : \_ \rightleftharpoons \_\mathrm{e} / X^{+}C\_\#$
MORDRE	mɔʁd	mɔʁde	
LEVER	lɛv	ləve	$\pi_2 : \_ \_ \rightleftharpoons \_\mathrm{ə}\_\mathrm{e} / X^{+}\_C\_\#$
MENER	mɛn	məne	
CROIRE	kʁwa	kwaje	$\pi_3 : \_ \rightleftharpoons \_\mathrm{je} / X^{+}\mathrm{wa}\_\#$

2. Classify predictor cell shapes on the basis of which patterns they are compatible with.

Form predictability as average surprisal v

Lexeme	PRS.3PL	PRS.2PL	π_1	π_2	π_3	Predictor shape
BAVER	bav	bave	✓			κ_1
MORDRE	mɔʁd	mɔʁde	✓			
PEINER	pɛn	pɛne	✓	✓		κ_2
LEVER	lɛv	ləve	✓	✓		
MENER	mɛn	məne	✓	✓		
CROIRE	kʁwa	kwaje			✓	κ_3

$\Rightarrow$ Puts words from predictor cell c into classes $\kappa_1, \ldots, \kappa_m$ that share phonological properties relevant for determining what happens in cell c'.

Form predictability as average surprisal vi

Lexeme	PRS.3PL	PRS.2PL	π_1	π_2	π_3	Predictor shape
BAVER	bav	bave	✓			κ_1
MORDRE	mɔʁd	mɔʁde	✓			
PEINER	pɛn	pɛne	✓	✓		κ_2
LEVER	lɛv	ləve	✓	✓		
MENER	mɛn	məne	✓	✓		
CROIRE	kʁwa	kwaje			✓	κ_3

3. Compute the **surprisal of the form found in cell c' given the form found in cell c**:

$$S = -\log_2 P(\pi_i \mid \kappa_j)$$

Form predictability as average surprisal vii

Lexeme	PRS.3PL	PRS.2PL	Pattern	Class	p	S
BAVER	bav	bave	π_1	κ_1	1	0
MORDRE	mɔʁd	mɔʁde	π_1	κ_1	1	0
PEINER	pɛn	pɛne	π_1	κ_2	13	1.585
LEVER	lɛv	ləve	π_2	κ_2	23	0.585
MENER	mɛn	məne	π_2	κ_2	23	0.585
CROIRE	kʁwa	kwaje	π_3	κ_3	1	0

Form predictability as average surprisal viii

4. Average over predictor cells c to get an overall estimation of how surprising c' is given the rest of the paradigm.

    - Ideally, this should be weighted by cell frequency.
    - But we do not have quality estimations of cell frequency, because of pervasive syncretism.
    - For lack of a better solution we use unweighted frequency.

Form predictability as local entropy

- Instead of asking how surprising the actual form is, we can ask how much uncertainty is associated with the distribution of possible forms.
- To that effect we can use the entropy of the distribution of patterns sharing a class, which we call **local entropy**. For class κ_j:

$$H = -\sum_{\pi \in \Pi} P(\pi \mid \kappa_j) \times \log_2 P(\pi \mid \kappa_j)$$

Lexeme	PRS.3PL	PRS.2PL	π_1	π_2	π_3	p	S	H
BAVER	bav	bave	✓			1	0	0
MOURIR	mɔʁd	mɔʁde	✓			1	0	0
PEINER	pɛn	pɛne	✓	✓		13	1.585	0.918
LEVER	lɛv	ləve	✓	✓		23	0.585	0.918
MENER	mɛn	məne	✓	✓		23	0.585	0.918