Stem spaces and regularity in verbal inflection

Olivier Bonami
Université Paris-Sorbonne / LLF
olivier.bonami@paris-sorbonne.fr

Fabio Montermini
CLLE-ERSS, CNRS & Université de Toulouse
fabio.montermini@univ-tlse2.fr

Understanding the Architecture of the Mental Lexicon
Pisa, 24-26 November 2011

Introduction

- An overview of some recent research on inflection in Romance languages
- A reassessment of the developments morphology has known in the last 20 years
- Morphomes, stem spaces as means of structuring paradigms
- Integrating cognitively plausible models with formal linguistic treatments

Introduction

- Structuring paradigms
- Implications in paradigms
- Stem spaces and entropy
- Stem spaces and graphs
- Conclusions

Structuring paradigms

- Structuralist (and many generative) models of morphology considered formal unity as a principle: (almost) all formal variation is reduced by posing (morpho)phonological rules modifying a basic form.
- However, we the linguists know the very reason why this variation is observed, since it is the reflect of historical changes the language has gone through.
- A consequence of this was the projection in syncrony of historical changes. However, to our knowledge, no serious discussion of the cognitive premises and consequences of this has never been put forth.

Structuring paradigms

- Such notions as morphomes (Aronoff 1994, Maiden 2005, 2009) are a way of structuring paradigms according to purely morphological principles; i.e. independent of external (phonological, semantic) motivations.

Structuring paradigms

- Morphomic distribution of verbal forms in Nepali

Long negative present of BIRSANU ('forget')

	M.SG	F.SG	PL
1	birsãdinã	birsãdinã	birsãdajnaũ
2.LOW	birsãdajnas	birsãdinas	birsãdajnau
2.MID	birsãdajnau	birsãdinau	birsãdajnau
3.LOW	birsãdajna	birsãdina	birsãdajnan
3.MID	birsãdajnan	birsãdinan	birsãdajnan
HIGH	birsanuhũdajna		

3 grades of honorifics for persons 2-3; high grade honorific neutralizes person-gender-number (Bonami & Boyé 2010)

Structuring paradigms

- Morphomic distribution of verbal forms in Nepali

  Long negative present of BIRSANU ('forget')

	M.SG	F.SG	PL
1	birsãd-**i**-**na**-**ã**	birsãd-**i**-**na**-**ã**	birsãd-**aj**-**na**-**aũ**
2.LOW	birsãd-**aj**-**na**-**s**	birsãd-**i**-**na**-**s**	birsãd-**aj**-**na**-**au**
2.MID	birsãd-**aj**-**na**-**au**	birsãd-**i**-**na**-**au**	birsãd-**aj**-**na**-**au**
3.LOW	birsãd-**aj**-**na**	birsãd-**i**-**na**	birsãd-**aj**-**na**-**n**
3.MID	birsãd-**aj**-**na**-**n**	birsãd-**i**-**na**-**n**	birsãd-**aj**-**na**-**n**

Structuring paradigms

- All the inflected forms are considered as equal: what counts in a paradigm are global relations between all the forms, not local relations between a form and the hypothetical basic form of a lexeme it derives from.
- Such structures as morphomes have proven to be useful to account for:
  - diachronic change (Maiden 2005, 2009)
  - psycho-computational models (Pirrelli 2000; Bonami *et al.* 2008)

Structuring paradigms

- The most neutral way of addressing the question of paradigmatic relations is the "Paradigm Cell Filling Problem" (Ackerman *et al.* 2009; Malouf & Ackerman 2010):

  "Given exposure to an inflected wordform of a novel lexeme, what licenses reliable inferences about the other wordforms in its inflectional family?"

Structuring paradigms

- Knowledge of implicative patterns relating cells in a paradigm is relevant.
  - This knowledge is best characterized in information-theoretic terms
  - The reliability of implicative patterns relating paradigm cell A to paradigm cell B is measured by the conditional entropy of cell B knowing A.

  (Ackerman *et al.* 2009; Malouf & Ackerman 2010)

Structuring paradigms

- In theory, a solution to the PCFP should consist in determining how any word-form in a paradigm can be deduced from any other form, i.e. in describing the relations between all the forms of a lexeme (e.g. for a French verb 51 × 50 = 2,550 relations).
- Probably, such a distributed view is the most plausible model of morphological competence from a cognitive perspective.
- However, what linguists should be interested in is describing and formalizing languages and the differences between them.
- The model we propose aims at accounting for higher level properties of such a system.

Structuring paradigms

- A simple inventory of all form-to-form relations in a paradigm is a huge task.
- Moreover, it is probable that it hides some generalizations.
- In fact, morphologists have elaborated several theoretical objects aiming at giving structure to paradigms, intuitively seen as networks of relations between equal forms; morphomes are one of these objects.

Structuring paradigms

- In theory, all the relations between forms in a paradigm should be considerer as equal. In practice, apart from cases of extremely simple inflectional systems, languages do not function like this.
- Several phenomena (allomorphy, syncretism, inflectional class repartition...) influence the degree of predictability of each relation.

Implications in paradigms

- Measuring the existence of structure in implicative relations.
- When paradigms are small, the structure is transparent

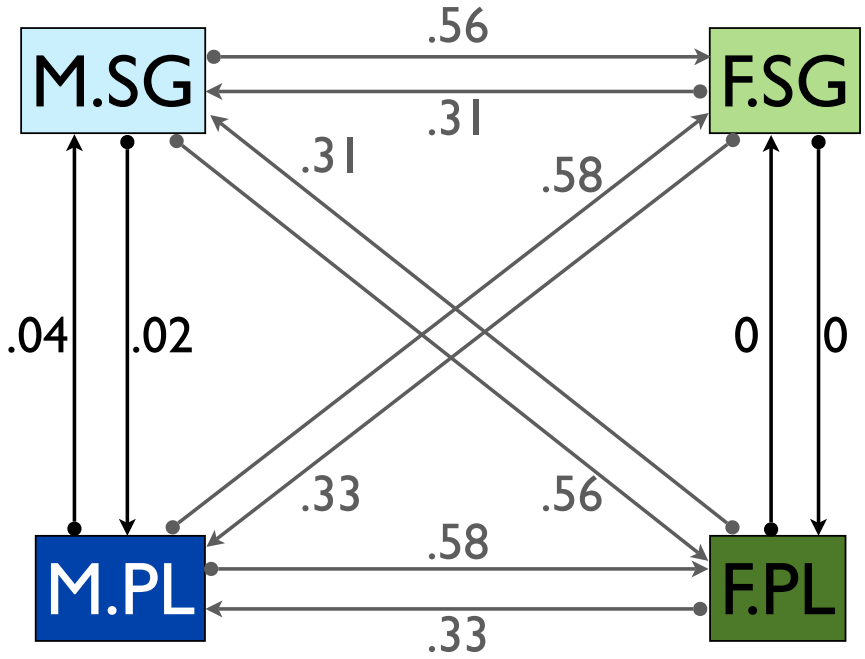
French adjectives

	M SG	M PL	F SG	F PL	
'hard'	dyʁ	dyʁ	dyʁ	dyʁ	77%
'legal'	legal	lego	legal	legal	3%
strong'	fɔʁ	fɔʁ	fɔʁt	fɔʁt	20%

Approximative figures from the `lexique.org` database

Implications in paradigms

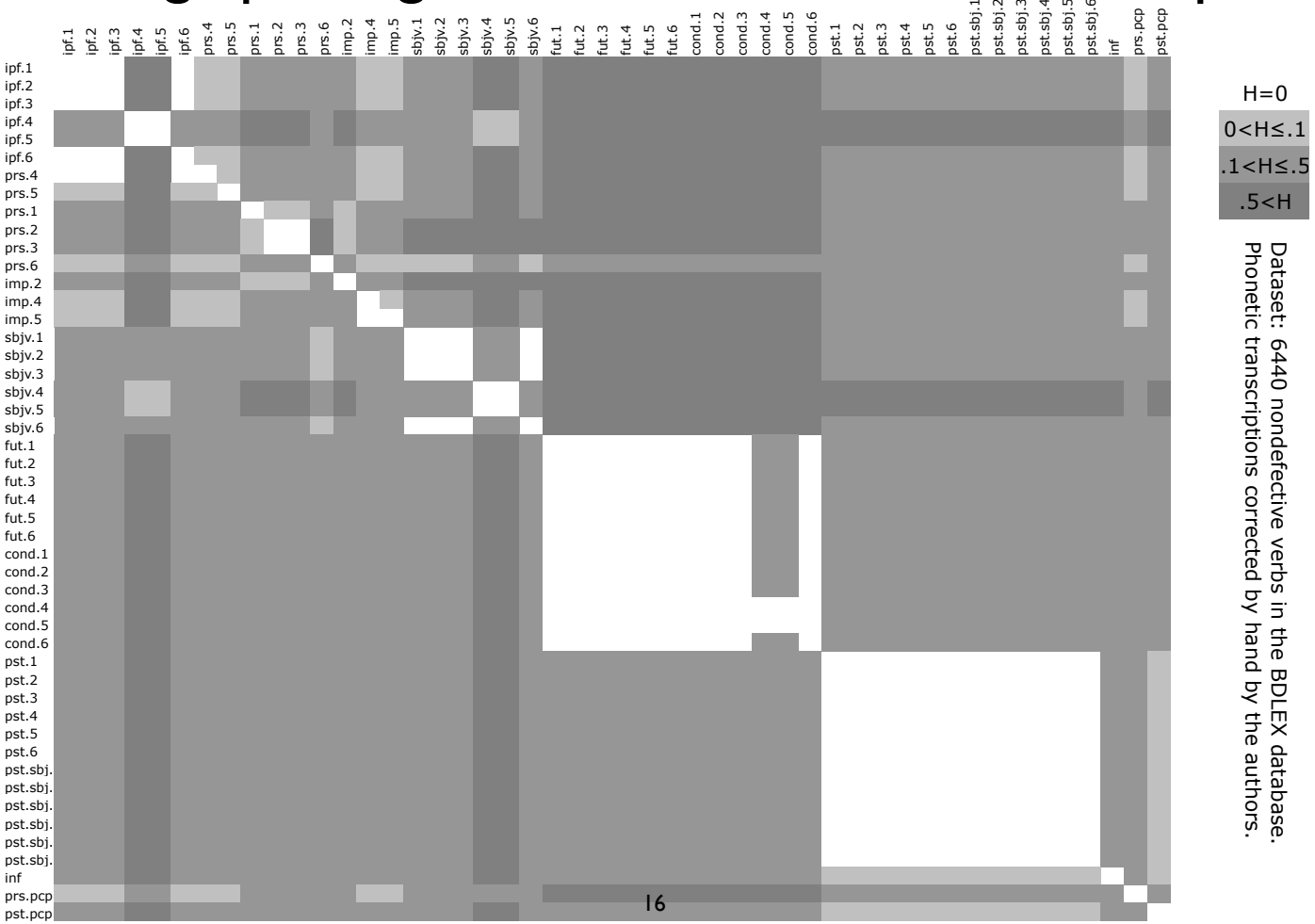
- Entropy measures of predictability between the forms of French adjectives



Dataset: 4,972 fully inflected adjectives from the `lexique.org` database. Phonetic transcriptions corrected by hand by the authors[5]

Implications in paradigms

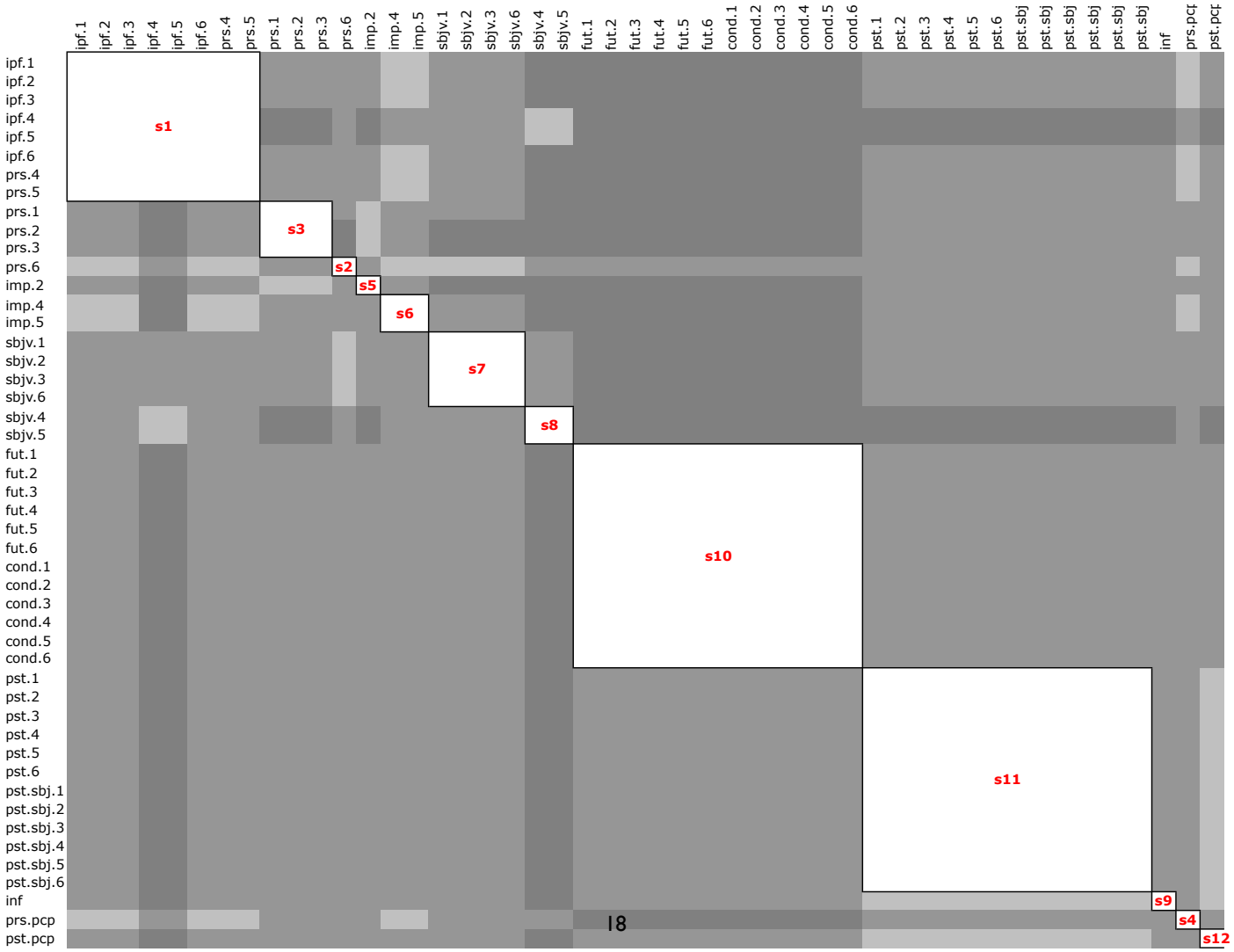
- With large paradigms, measures become uninterpretable



Stem spaces and entropy

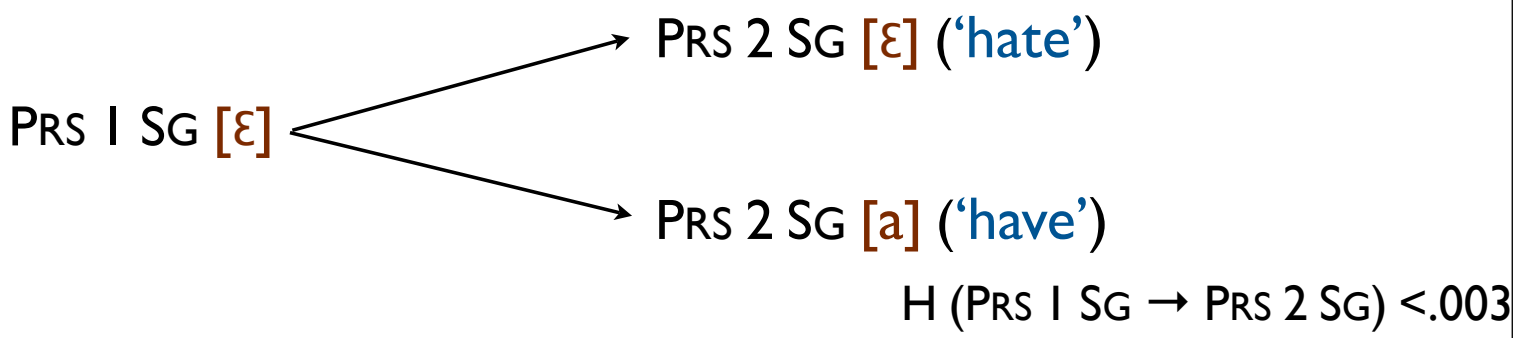
- Stem spaces for French (Bonami & Boyé 2003, 2007, Boyé 2009), Italian (Pirrelli & Battista 2000, Montermini & Boyé f.c.), Spanish (Boyé & Cabredo-Hofherr 2006), Catalan (Guerrero f.c.) were designed by hand as a way of capturing qualitatively implicative structure in paradigms.
- It turns out that there is a strong correlation between the hand-designed stem spaces and quantitative properties of implicational relations.
- In the case of French, slots in the stem space correspond almost exactly to the collections of paradigm cells related by fully predictable implicatonal relations.

Stem spaces and entropy

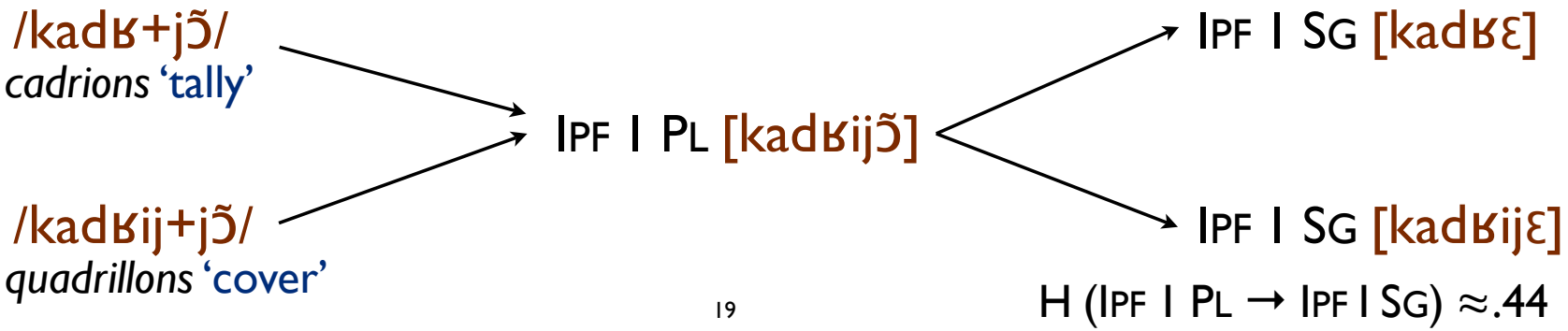


Stem spaces and entropy

- Two types of exceptions to this observation
  - local uncertainities introduced by a single ambiguous pair, giving rise to very low entropy (<.01)



  - phonological opacity



Stem spaces and graphs

- The notion of a stem space is based on the observation that, in a paradigm, some relations are more predictive than others.
- There are clusters of forms that are in systematic covariation, i.e. they systematically display the maximum degree of interpredictability.
- Conversely, the relations between these clusters may or may not display a high degree of interpredictability.
- Paradigms are considered as organised into stem spaces, a pattern defining the (morphomic) distribution of stems among word-forms.

Stem spaces and graphs

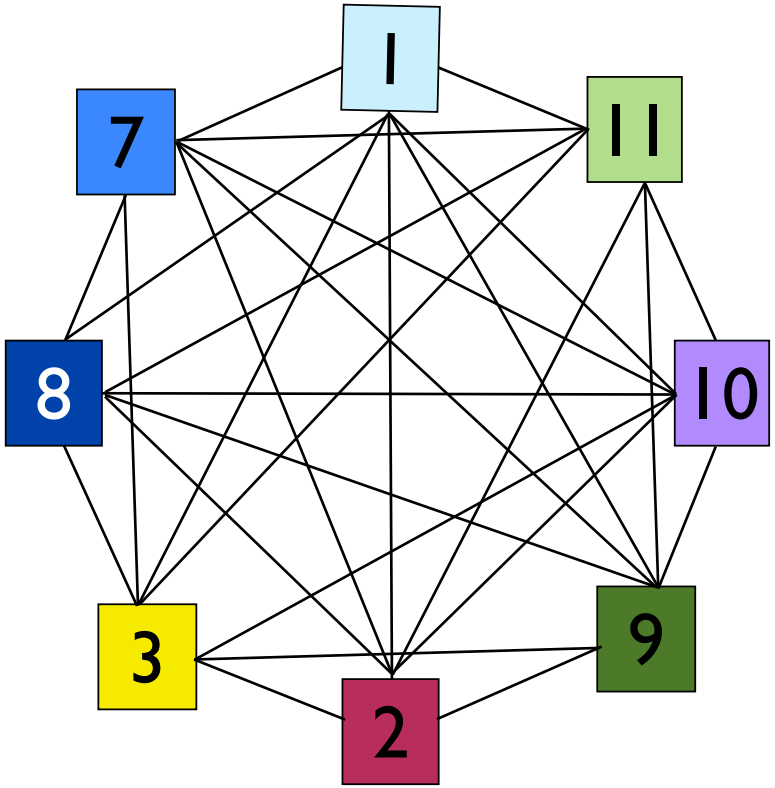
	1SG	2SG	3SG	1PL	2PL	3PL
PRES IND	3	3	3	1	1	2
IMPF IND	1	1	1	1	1	1
PRES SBJ	7	7	7	8	8	7
FUT IND	10	10	10	10	10	10
PRES CND	10	10	10	10	10	10
PAST IND	11	11	11	11	11	11
IMPF SBJ	11	11	11	11	11	11
IMPER		5		6	6	

INFIN	PRES PART	PAST PART			
		M SG	F SG	M PL	F PL
9	4	12	12	12	12

Stem space of French verbs

Stem spaces and graphs

- In theory, 56 relations are relevant for filling a 8-stem space:



Stem spaces and graphs

- French verb forms

1 PL PRES IND (S1)	3 PL PRES IND (S2)	1 SG PRES IND (S3)	
lavɔ̃	lav	lav	‘wash’
muʁɔ̃	mœʁ	mœʁ	‘die’
savɔ̃	sav	sɛ	‘know’
byvɔ̃	bwav	bwa	‘drink’

Stem spaces and graphs

- French verb forms

1 PL PRES IND (S1)	3 PL PRES IND (S2)	1 SG PRES IND (S3)
A	A	A
A	B	B
A	A	B
A	B	C

Stem spaces and graphs

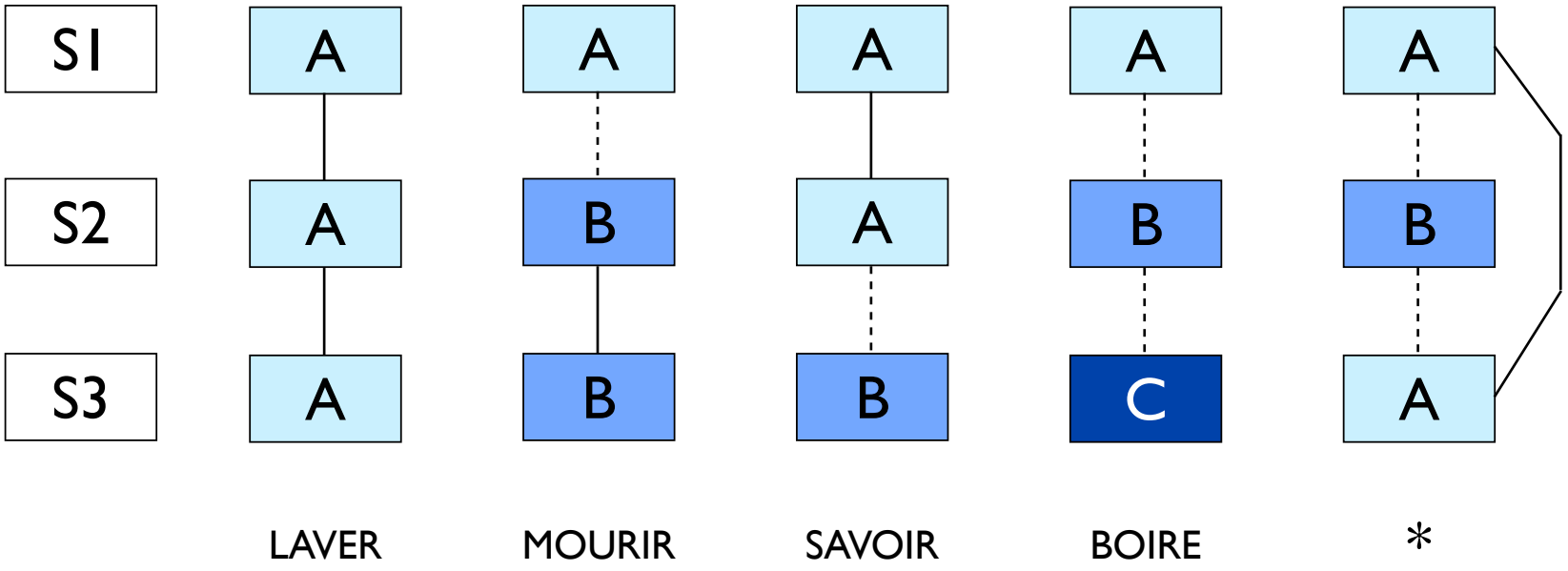
- French verb forms

1 PL PRES IND (S1)	3 PL PRES IND (S2)	1 SG PRES IND (S3)
A	A	A
A	B	B
A	A	B
A	B	C
A	B	A

Not attested

Stem spaces and graphs

- Knowing S1 in addition to S2 does not improve the chances of predicting S3



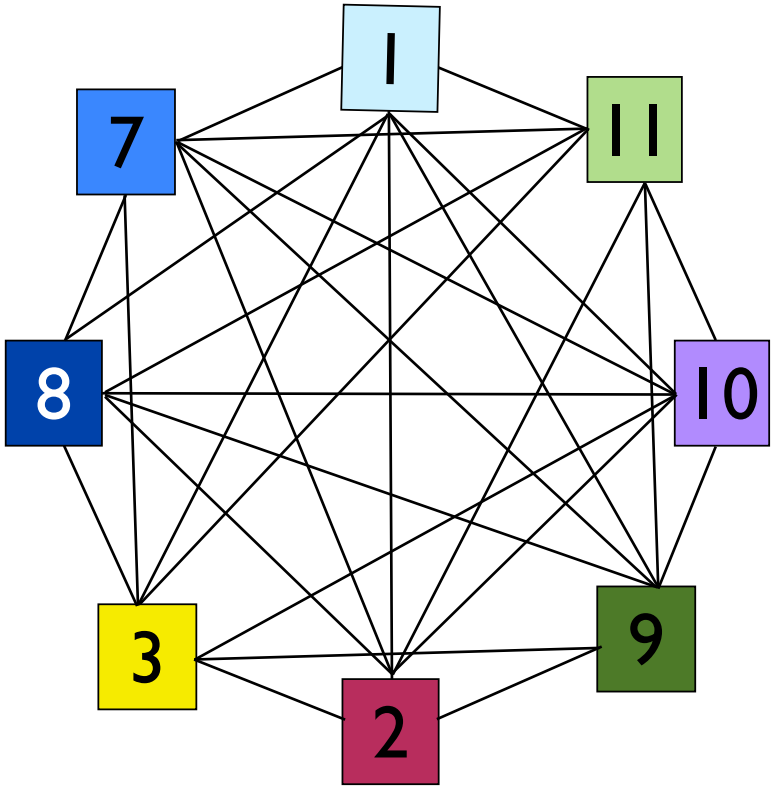
Stem spaces and graphs

- In fact, most of the relations display a low degree of predictability

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12
s1		0,23	0,47	0,00	0,47	0,01	0,23	0,68	0,39	0,59	0,40	0,37
s2	0,09		0,29	0,09	0,29	0,10	0,01	0,45	0,43	0,43	0,43	0,43
s3	0,35	0,50		0,33	0,00	0,33	0,51	0,67	0,40	0,71	0,37	0,36
s4	0,00	0,23	0,48		0,48	0,01	0,23	0,28	0,39	0,59	0,40	0,37
s5	0,35	0,49	0,00	0,34		0,35	0,52	0,69	0,40	0,71	0,37	0,36
s6	0,00	0,20	0,47	0,00	0,47		0,23	0,68	0,39	0,59	0,39	0,37
s7	0,10	0,02	0,28	0,10	0,28	0,11		0,29	0,42	0,56	0,42	0,41
s8	0,42	0,50	0,68	0,44	0,68	0,44	0,49		0,76	0,86	0,72	0,68
s9	0,13	0,30	0,23	0,13	0,23	0,13	0,30	0,76		0,29	0,08	0,07
s10	0,31	0,29	0,11	0,31	0,11	0,31	0,29	0,72	0,28		0,30	0,28
s11	0,18	0,36	0,29	0,18	0,29	0,18	0,36	0,83	0,13	0,40		0,07
s12	0,17	0,36	0,26	0,17	0,26	0,17	0,35	0,80	0,11	0,36	0,07	

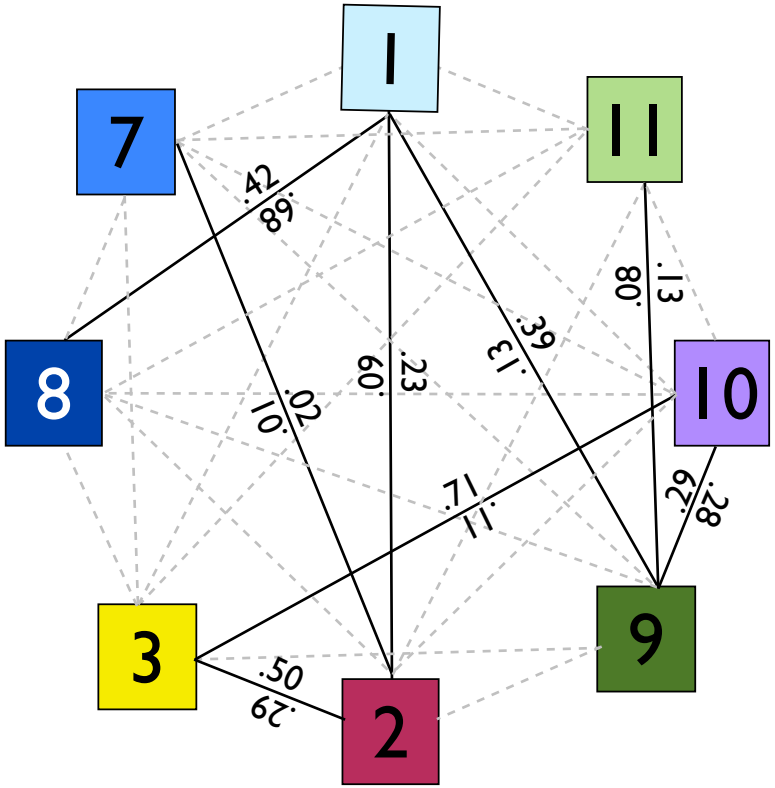
Stem spaces and graphs

- In fact, most of the relations display a low degree of predictability



Stem spaces and graphs

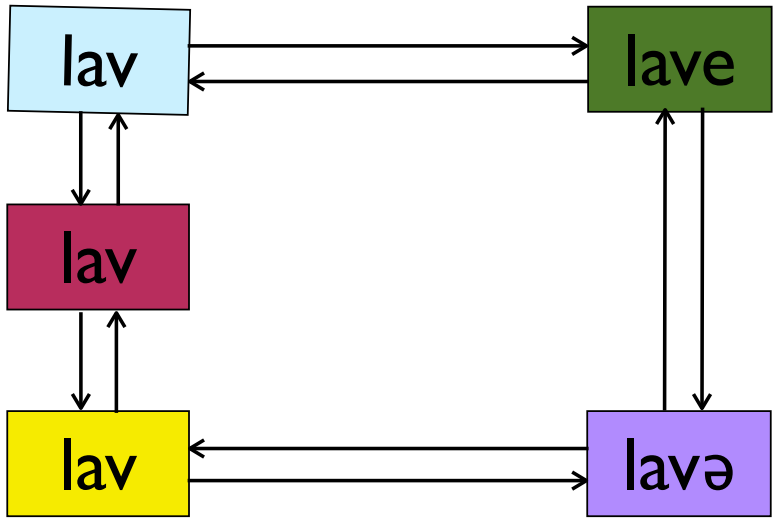
- In fact, most of the relations display a low degree of predictability



Stem spaces and graphs

- Taking into account the relations allows drawing dependency graphs of paradigmatic relations

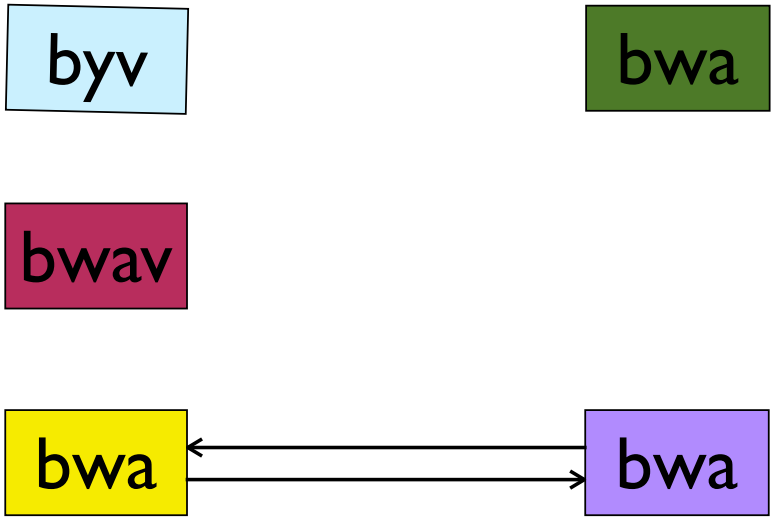
Graph of relations of LAVER ('wash')



Stem spaces and graphs

- Taking into account the relations allows drawing dependency graphs of paradigmatic relations

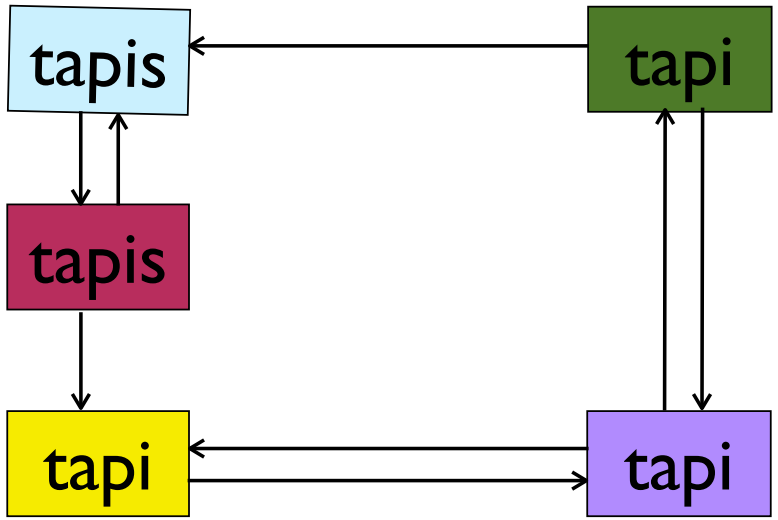
Graph of relations of BOIRE ('drink')



Stem spaces and graphs

- Taking into account the relations allows drawing dependency graphs of paradigmatic relations

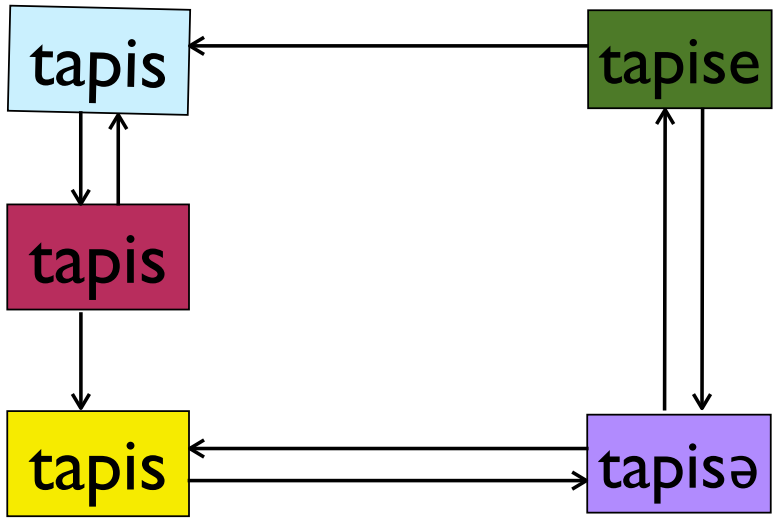
Graph of relations of TAPIR ('cover')



Stem spaces and graphs

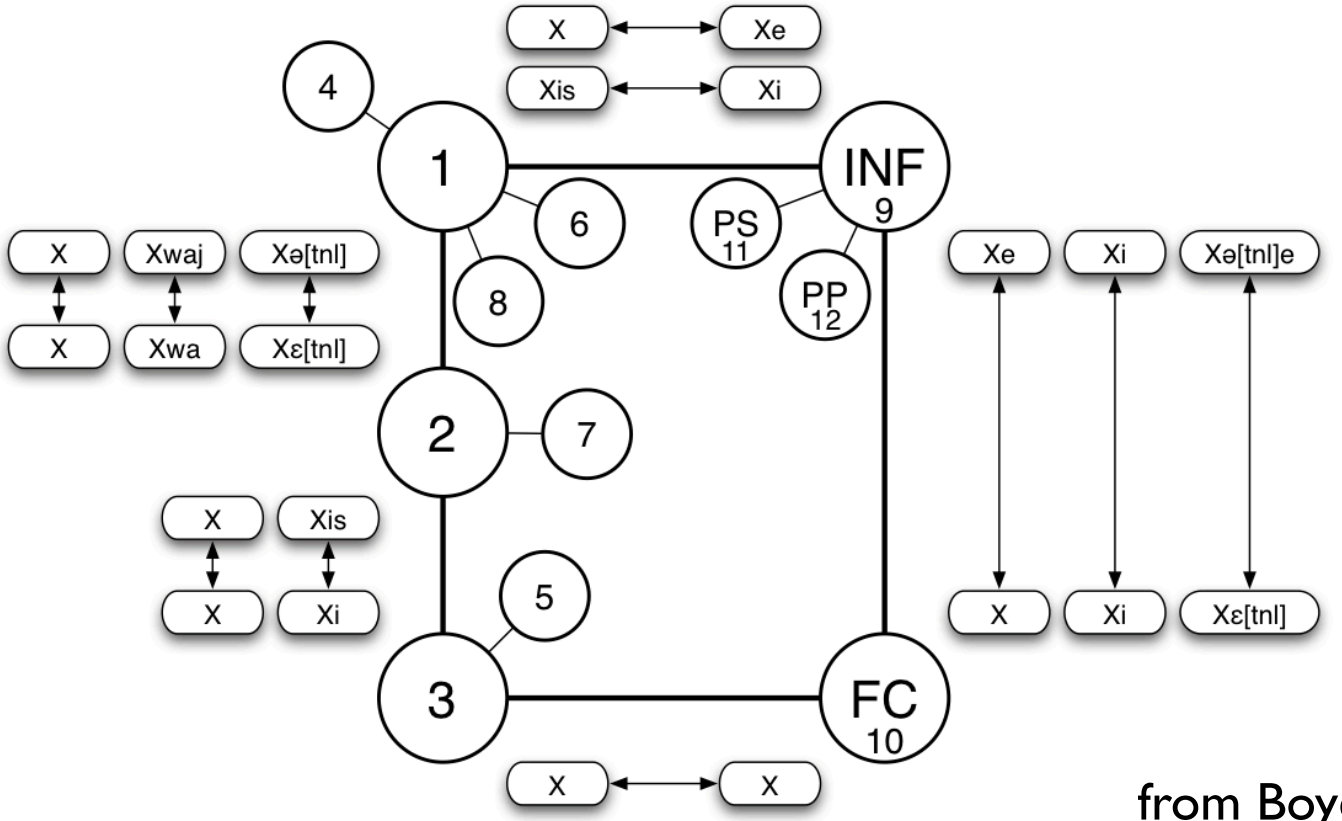
- Taking into account the relations allows drawing dependency graphs of paradigmatic relations

Graph of relations of TAPISSER ('paper')



Stem spaces and graphs

- Full graph of relations for regular verbs in French



from Boyé (2010)

Conclusion

- The autonomy of morphology (of purely morphological relations) poses the problem of reducing the complexity of inflectional systems by relying only on morphological relations.
- Although fully connectionist (surface-to-surface, word-based) models may be cognitively plausible, we need more in order to account for the variety of languages and of phenomena (recurrent patterns of allomorphy / syncretism; inflectional classes and subclasses...).
- Stem spaces and graphs (among others) are means of reducing the complexity of computation needed in inflectional morphology while maintaining the claim that lexemes are complex objects.
- Formally, they are best represented as networks of connected forms rather than unique phonological objects that undergo (morpho)phonological operations.

References

Ackerman F., Blevins J.P., Malouf R. 2009: "Parts and wholes: implicative structures in inflectional paradigms" *in* J. Blevins, J.P. Blevins (eds), *Analogy in Grammar*. Oxford: Oxford University Press, 54-82.

Aronoff M. 1994: *Morphology by Itself*. Cambridge, MA: MIT Press.

Blevins J.P 2006: "Word-based morphology" *Journal of Linguistics* 42: 531-573.

Bonami O., Boyé G. 2003: "Supplétion et classes flexionnelles en français". *Langages* 152: 102-126.

Bonami O., Boyé G. 2007: "Remarques sur les bases de la conjugaison" *in* E. Delais-Roussarie, L. Labrune (eds), *Des sons et des sens. Modèles et données en morphologie et en phonologie*. Paris: Hermès-Lavoisier: 77-90.

Bonami O., Boyé G. 2010: "Opaque paradigms, transparent forms in Nepali conjugation". *WOTM5*, June 2010.

Bonami O., Boyé G., Giraudo H. Voga M. 2008: "Quels verbes sont réguliers en français?" *in* J. Durand, B. Habert, B. Laks (eds), *CMLF08*. Paris: Institut de Linguistique Française, 1511-1523.

Boyé G. 2010: "Régularités et classes flexionnelles dans la conjugaison du français" *in* M. Roché, G. Boyé, N. Hathout, S. Lignon, M. Plénat *Des unités miorphologiques au lexique*. Paris: Hermès-Lavoisier, 41-68.

References

Boyé G., Cabredo-Hofherr P. 2006: "The structure of allomorphy in Spanish verbal inflection". *Cuadernos de Lingüística del Instituto Ortega y Gasset* 13: 9-24.

Guerrero A. f.c. "Verbal inflection in Central Catalan: A realizational analysis". *Lingue e Linguaggio*.

Maiden M. 2005: "Morphological autonomy and diachrony". *Yearbook of Morphology 2004*: 137-175.

Maiden M. 2009: "From pure phonology to pure morphology. The reshaping of the Romance verb". *Recherches de linguistique de Vincennes* 38: 45-82.

Malouf R., Ackerman F. 2010: "Paradigms: the low entropy conjecture". Paper presented at the Workshop on Morphology and Formal Grammar, Paris.

Montermini F., Boyé G. f.c. :"Stem relations and inflection class assignment in Italian". *Word Structure*.

Pirrelli V. 2000: *Paradigmi in morfologia. Un approccio interdisciplinare alla flessione verbale dell'italiano*. Roma: Istituti Editoriali e Poligrafici Internazionali.

Pirrelli V., Battista M. 2000: "The paradigmatic dimension of stem allomorphy in Italian". *Rivista di linguistica* 12.2: 307-380.