

# SUR LA MORPHOLOGIE IMPLICATIVE DANS LA CONJUGAISON DU PORTUGAIS : UNE ETUDE QUANTITATIVE

**Olivier Bonami**  
Université Paris-Sorbonne  
Institut Universitaire de France  
Laboratoire de Linguistique Formelle

**Ana R. Luís**  
Universidade de Coimbra  
CELGA

À paraître dans les *Mémoires de la Société de linguistique de Paris*, vol. 22

## Abstract

Adopting a word-based approach to verbal inflection, this paper proposes an automatic study of the conjugation system of European Portuguese with the two-fold goal of identifying patterns of alternation between forms within the paradigm and quantifying interpredictability between paradigm cells. Based on a newly developed lexicon of 1000 non-defective verbs (corresponding to 68 000 inflected forms), given in broad phonetic transcription, we apply tools developed by Bonami&Boyé (in press) for French, in line with (one variant of) the information-based approach of Ackerman, Blevins & Malouf (2009). Through the systematic application of this research methodology to the phoneticised lexicon of European Portuguese, this study reveals that the system of European Portuguese theme vowels, which is very similar to that of Latin, interacts with phenomena such as vowel reduction giving rise to very specific patterns of opacity. By applying a methodology originally developed for French to a closely related language such as European Portuguese, this study also reveals that while both languages share similar phenomena (e.g., root allomorphy, vowel reductions, theme vowels), the different proportions with which such phenomena occur reveal the existence of two distinct inflectional profiles.

## I. Introduction

Peu de linguistes considèrent l'étude des propriétés formelles des systèmes flexionnels synchroniques de langues bien documentées comme une tâche particulièrement excitante<sup>1</sup>. Après tout, ces systèmes sont réputés être décrits en détail par les grammaires pédagogiques, qui véhiculent l'idée que le système est constitué de quelques classes régulières et d'un catalogue informe d'irrégularités. Le présent article vise à montrer par l'exemple que le sujet mérite l'attention, et ce pour au moins trois raisons. D'abord, les descriptions des grammaires pédagogiques sont presque toujours centrées sur l'écrit ; or les conventions orthographiques tendent à donner une image fautive de la complexité du système, que ce soit pour la diminuer (par exemple en désambiguïsant des neutralisations phonologiques) ou pour l'augmenter (par exemple en introduisant des distinctions orthographiques sans réflexe phonétique). Ensuite,

---

<sup>1</sup> Les travaux présentés ici ont été partiellement financés par le programme PHC PESSOA (Campus). Nous remercions chaleureusement quatre collègues sans lesquels ces travaux n'auraient pu être menés à bien : Gilles Boyé (U. Bordeaux 3) et Delphine Tribout (U. Lille 3), qui ont contribué à l'écriture des outils informatiques utilisés ; Sara Candeias et Fernando Perdigão (U. de Coimbra), qui ont mis à notre disposition leur dictionnaire de prononciation et nous ont aidés à l'adapter aux besoins du présent projet. Nous remercions également les participants de la journée d'étude 2013 de la SLP pour leurs critiques, commentaires et suggestions, et plus particulièrement Vittorio dell'Aquila, Aurélie Guerrero, Jean-Léo Léonard, Michele Loporcaro, et Martin Maiden. Des discussions avec James Blevins et Yves-Charles Morin ont été cruciales pour retracer l'histoire des travaux sur la morphologie implicative.

les descriptions traditionnelles ne permettent généralement pas de connaître l'ampleur des phénomènes observés. Il existe par exemple bien peu de langues pour lesquelles on dispose de guides de conjugaison permettant d'établir comment différentes classes flexionnelles sont peuplées ; même quand c'est le cas, ces guides ne s'attachent généralement pas à faire ressortir les méta-régularités qui structurent le système dans son ensemble. Enfin, les méthodes d'analyse nouvelles gagnent à être appliquées à des systèmes bien connus et bien compris, de manière à ce que la critique soit aisée sur la base des descriptions existantes. Même si on a pour objectif de long terme d'étudier des langues typologiquement variées et peu documentées, la validation de la méthode sur des langues familières est donc un préalable méthodologique utile.

Cet article présente une étude semi-automatique du système de conjugaison du portugais européen. Ce système a deux caractéristiques qui le rendent particulièrement intéressant de ce point de vue : d'une part il est relativement mal connu, en comparaison aux conjugaisons des autres langues romanes majeures ; d'autre part, la phonologie du portugais a des caractéristiques remarquables qui ont une influence indirecte sur sa morphologie et lui donnent une place à part dans le domaine roman. L'analyse est basée sur un lexique de 1000 verbes fréquents non-défectifs, soit 68 000 formes fléchies, en transcription phonétique ; ce système a été analysé automatiquement en utilisant les outils développés par Bonami & Boyé (sous presse), qui implémentent une variante de l'approche informationnelle de la morphologie défendue par Ackerman, Blevins & Malouf (2009). Sur la base de cette préanalyse automatique, nous montrons comment le système de voyelles thématiques du portugais, très proche de celui du latin, interagit avec les phénomènes de réduction vocalique pour donner lieu à des situations d'opacité très caractéristiques. Globalement, l'application d'une méthode initialement conçue pour le français à une langue proche permet de faire ressortir l'existence de différences importantes entre les deux systèmes. Nous soutenons que le caractère quantitatif de l'approche, et la systématisme autorisée par l'automatisation, sont cruciales pour arriver à ce résultat : les deux langues connaissent des phénomènes similaires (allomorphie du radical, réductions vocaliques, voyelles thématiques) mais dans des proportions si variables qu'elles produisent un profil général remarquablement différent.

L'article est structuré de la manière suivante : dans la deuxième partie, nous esquissons le contexte intellectuel dans lequel cette étude s'est développée, qui est celui des études sur les relations d'interprédictibilité dans les conjugaisons des langues romanes. Dans la troisième partie, nous présentons l'ensemble de données utilisé et résumons la procédure d'analyse implémentée qui a été utilisée. La quatrième partie présente les résultats de la préanalyse automatique, et identifie les principales causes d'imprédictibilité que celle-ci permet de mettre au jour. La cinquième et dernière partie contraste le profil de la conjugaison du portugais européen qui résulte de cette étude avec celui établi par Bonami & Boyé (sous presse) pour le français.

## **II. De l'allomorphie radicale à la morphologie implicative**

Au cours des vingt-cinq dernières années, l'étude des systèmes flexionnels romans a occupé le centre de l'attention des théoriciens de la flexion, sous l'impulsion des travaux d'Aronoff (1994), reprenant et développant les conséquences des observations de Matthews (1972) sur les « formations parasitiques », et de Maiden (1992) sur le développement diachronique des patrons de conjugaison romans. La motivation théorique de cette focalisation nouvelle était la réflexion sur la présence dans les paradigmes de flexion de régularités purement morphologiques, c'est à dire qui ne peuvent être expliquées ni par l'expression d'une même propriété syntaxique ou sémantique, ni par un conditionnement

phonologique homogène. Les phénomènes d'allomorphie et de supplétion radicales dans la conjugaison du français constituent un exemple frappant de cette situation. Si on examine le paradigme du verbe ALLER, illustré dans le tableau 1, il apparaît clairement que le choix du radical est parfois motivé (le futur et le conditionnel, qui partagent le radical /i-/, constituant vraisemblablement une classe naturelle, cf. par exemple Verkuyl *et al.* 2004), et parfois arbitraire : le choix d'un radical en /v-/, /al-/, ou /aj-/ découpe le paradigme en trois sections dont aucune ne correspond à une classe naturelle, que ce soit en termes syntaxiques, sémantiques ou phonologiques.

	SG			PL		
	1	2	3	1	2	3
PRESENT	vɛ	va	va	alʃ	ale	vɔ̃
IMPARFAIT	alɛ	alɛ	alɛ	aljɔ̃	alje	alɛ
PASSE SIMPLE	alɛ	ala	ala	alam	alat	alɛʁ
SUBJONCTIF IMPARFAIT	alas	alas	ala	alajsɔ̃	alasje	alas
SUBJONCTIF PRESENT	aj	aj	aj	aljɔ̃	alje	aj
FUTUR	iʁɛ	iʁa	iʁa	iʁɔ̃	iʁe	iʁɔ̃
CONDITIONNEL	iʁɛ	iʁɛ	iʁɛ	iʁjɔ̃	iʁje	iʁɛ

INFINITIF	PARTICIPE PRESENT	PARTICIPE PASSE			
		M.SG	F.SG	M.PL	F.PL
aller	allé	allé	allée	allés	allées

**Tableau 1** : La flexion du verbe ALLER

L'inventaire de ce type de situation dans le domaine roman, ainsi que la réflexion sur leur modélisation, a donné lieu à de nombreux travaux fructueux, parmi lesquels on peut notamment citer Pirrelli & Battista (2000) sur l'italien, Bonami & Boyé (2003) pour le français, Boyé & Cabredo Hofherr (2006) pour l'espagnol, Guerrero (2011) pour le catalan, Esher (2012) pour l'occitan. Parallèlement, les travaux de Maiden (notamment 2005, 2009) ont clairement établi la stabilité diachronique de tels patrons d'allomorphie immotivés—ou, pour reprendre le terme d'Aronoff (1994), morphomiques.

Les travaux cités jusqu'ici se focalisent sur la distribution des patrons d'allomorphie radicale, plutôt que sur leur structuration interne. Morin (1987) fut le premier à discuter l'existence de relations d'IMPLICATION entre les radicaux dans les paradigmes de conjugaison du français—c'est à dire de relations orientées de prédictibilité d'une forme à partir d'une autre. Le tableau 2 illustre les relations discutées par Morin, qui sont listées en (1). Dans l'interprétation de Morin, les relations d'implication entre radicaux sont des relations par défaut : elles rendent compte de régularités partielles qui sont infirmées par des cas de verbes irréguliers possédant un ou plusieurs radicaux allomorphiques. Ainsi les radicaux listés en gras dans le tableau 2 doivent être appris en tant que tels par les locuteurs. Ceux qui sont en italiques peuvent être déduits sur la base des relations d'implication : un locuteur compétent qui n'a jamais entendu le présent 1SG de LAVER mais qui connaît son présent 3PL peut tirer une inférence valide de sa connaissance du système du français ; une telle inférence ne sera pas valide dans le cas de SAVOIR.

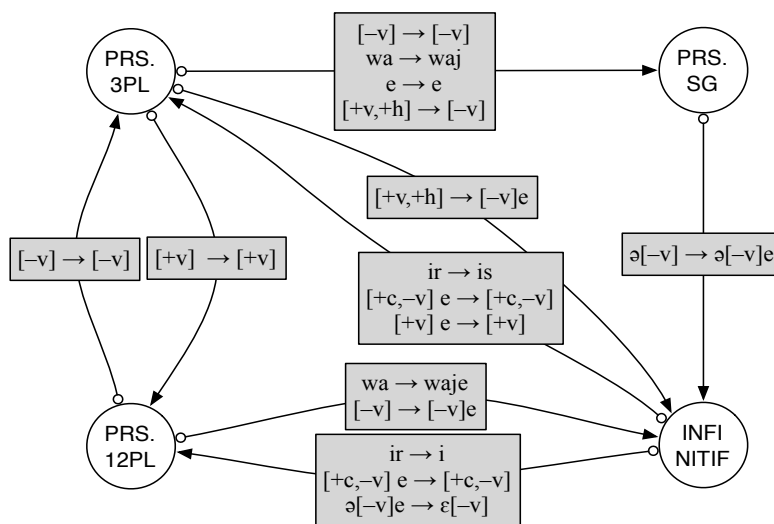
- (1) Relations implicatives au présent de l'indicatif en français (Morin 1987 : 37)
- a. Le radical de 1PL sert à la 2PL
  - b. Le radical de 1PL sert à la 3PL
  - c. Le radical de 3PL sert à la 3SG

- d. Le radical de 3SG sert à la 2SG
- e. Le radical de 2SG sert à la 1SG

	1PL	2PL	3PL	3SG	2SG	1SG
LAVÉ	<b>lavõ</b>	<i>lave</i>	<i>lav</i>	<i>lav</i>	<i>lav</i>	<i>lav</i>
SAVOIR	<b>savõ</b>	<i>save</i>	<i>sav</i>	<b>sɛ</b>	<i>sɛ</i>	<i>sɛ</i>
BOIRE	<b>byvõ</b>	<i>byve</i>	<b>bwav</b>	<b>bwa</b>	<i>bwa</i>	<i>bwa</i>
VOIR	<b>vwajõ</b>	<i>vwaje</i>	<b>vwa</b>	<i>vwa</i>	<i>vwa</i>	<i>vwa</i>
GOUAILLER	<b>gwajõ</b>	<i>gwaje</i>	<i>gwaj</i>	<i>gwaj</i>	<i>gwaj</i>	<i>gwaj</i>

**Tableau 2** : Quelques relations d'implication au présent de l'indicatif en français selon Morin (1987)

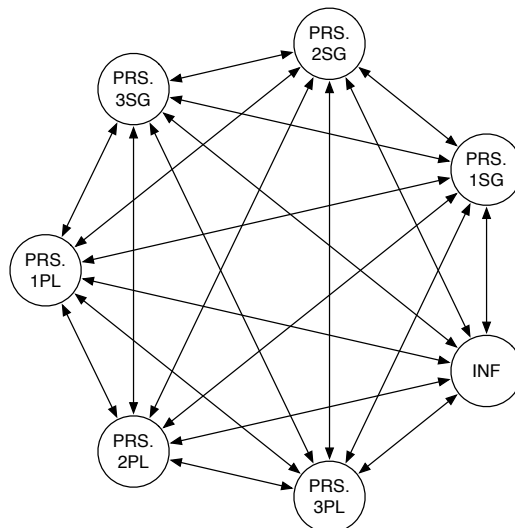
Les observations initiales de Morin ont été systématisées par Boyé (2000) et Bonami & Boyé (2003) sous la forme d'*Arbres de dépendances thématiques*. Boyé & Cabredo (2006) montrent sur la base de l'espagnol que le réseau des relations implicatives ne peut prendre la forme d'un arbre orienté, ce qui aurait pour conséquence qu'un locuteur connaissant par exemple une forme de présent singulier ne pourrait tirer d'inférence quant à aucune forme de présent pluriel. Ils substituent donc aux arbres des *graphes* de dépendances thématiques. Les gains descriptifs et théoriques apportés par les graphes sont explorés notamment dans Boyé (2011) pour le français, et Montermini & Boyé (2012) pour l'italien. Le graphe de la figure 1, basé sur Bonami & Boyé (2007), donne une image partielle des relations d'implication que l'on peut dès lors observer dans le sous-système de la conjugaison du français constitué par le présent de l'indicatif et l'infinitif. L'interprétation des relations d'implication est celle de Morin : chaque flèche indique une inférence non catégorique mais fiable qu'un locuteur est fondé à tirer s'il connaît la forme à l'origine de la flèche mais pas la forme de destination. La syntaxe des règles morphophonologiques est librement adaptée de celle de la phonologie générative classique (voir p. ex. Dell 1985).



**Figure 1** Relations implicatives entre radicaux en français selon Bonami & Boyé (2007)

Si les travaux sur les langues romanes se sont centrés sur la question de l'allomorphie radicale, l'examen des relations implicatives entre formes complètes, plutôt qu'entre radicaux, est lentement revenu au centre de l'attention pendant la même période. D'une part, Wurzel (1984) met l'examen de ces relations au centre de l'étude de la flexion, en se concentrant particulièrement sur l'interprétabilité entre désinences d'une case à l'autre du paradigme, que ce soit au sein d'une même classe flexionnelle ou à travers les classes.

D'autre part Zwicky (1985) puis Stump (2001) et Baerman, Brown & Corbett (2005) donnent une place centrale à la notion de *règle de renvoi* dans l'analyse du syncrétisme directionnel : de fait, ils argumentent que certains cas de syncrétisme doivent être modélisés en posant explicitement qu'une case du paradigme en parasite une autre. L'article fondateur de Blevins (2006) prend le contrepied de ces tentatives et place les relations implicatives au centre, plutôt qu'à la périphérie, des préoccupations du morphologue. Blevins montre que l'existence de régularités abstraites, comme des collections de formes partageant un radical, ou des segmentations stables en radical et affixe, n'est pas nécessaire à l'existence de relations d'interprédictibilité au sein des paradigmes de flexion, et que ces relations d'interprédictibilité doivent donc être considérées comme l'objet central de l'étude de la flexion. Autrement dit, dans une approche à la Blevins, l'objet d'étude du morphologue est le graphe complet des relations entre cases du paradigme, illustré dans la figure 2 en utilisant le même sous-ensemble des formes du français que précédemment. Dans ce contexte, un graphe comme celui de la figure 1 peut être conçu comme une abstraction qui peut être construite pour décrire de manière concise le graphe complet des relations d'interprédictibilité. Comme le soulignent Montermini & Bonami (2013), un graphe comme celui de la figure 1 fait ressortir les régularités présentes dans le système de relations d'implication entre formes, en mettant de côté les cas d'interprédictibilité mutuelle (quasi-)catégorique<sup>2</sup>.



**Figure 2** Relations implicatives à prendre en compte dans un modèle abstraitif

La discussion qui précède montre que la distance n'est pas grande entre une approche des relations implicatives centrée sur les alternances radicales, comme celle défendue par Boyé et ses collaborateurs à la suite de Morin, et une approche centrée sur les mots comme celle de Blevins, et, à sa suite, Ackerman, Blevins & Malouf (2009), Ackerman & Malouf (2013), Sims (2010), Bonami & Boyé (sous presse). Il existe cependant deux distinctions importantes entre les deux familles d'approches. D'une part, l'approche centrée sur les radicaux présuppose des hypothèses auxiliaires sur les propriétés morphologiques des systèmes étudiés ; en particulier elle présuppose que l'on puisse se mettre d'accord sur une procédure uniforme de segmentation. Or l'histoire de la morphologie, et en particulier de la morphologie

<sup>2</sup> Le graphe de la figure 1 fait abstraction de quelques cas isolés de formes fléchies irrégulières qui empêchent l'interprédictibilité d'être tout à fait totale au sein du présent singulier ou entre PRS.1PL et PRS.2PL. Voir Bonami & Boyé (2003) pour une motivation initiale et Bonami & Boyé (sous presse) pour une nouvelle discussion de ce problème du point de vue du cadre analytique adopté dans le présent article.

des langues romanes, montre que la segmentation est loin de suivre de l'observation directe des données ; voir Spencer (2012) pour une discussion récente. Dans l'approche basée sur les mots, aucune segmentation n'est nécessaire à l'observation des relations d'interprédictibilité, qui se base simplement sur l'observation des différences entre des couples de formes de surface. D'autre part, depuis Morin (1987), les défenseurs de l'approche centrée sur les radicaux s'appuient sur une notion de relation d'implication par défaut. Cela présuppose en retour qu'il existe une distinction catégorique entre patrons de flexion réguliers et irréguliers, le patron régulier étant celui qui sert de défaut. L'intention des chercheurs concernés est clairement que le choix du défaut devrait ultimement être motivé sur une base expérimentale ; l'expérimentation sur ce sujet étant cependant difficile (il y a plusieurs milliers de relations par défaut à observer pour les verbes français) et incertaine (il n'est pas clair qu'il existe toujours un patron par défaut ; voir par exemple Bonami *et al.* 2008), les travaux existants s'appuient de fait sur des observations impressionnistes. À l'inverse, l'approche centrée sur les mots s'est développée autour d'une conception probabiliste du choix des patrons d'alternance mis en jeu dans les relations d'interprédictibilité : la régularité n'est pas une propriété catégorique, et un patron est plus régulier qu'un autre s'il est plus probable qu'il soit choisi.

Cet article n'est pas le lieu pour une discussion des vertus conceptuelles (voir Blevins, à paraître) ou analytiques (voir Bonami & Boyé, sous presse) respectives des deux approches. En revanche il convient de noter que les propriétés de l'approche centrée sur les mots lui donnent un caractère opératoire distinctif : les patrons d'alternance entre formes de surface peuvent être observés directement, et de manière (semi-)automatique, à partir d'un lexique de ces formes ; la probabilité d'utilisation d'un patron peut elle aussi être approximée à partir de données de fréquence issues d'un lexique et/ou d'un corpus. De ce fait, l'approche centrée sur les mots ouvre la perspective d'une étude semi-automatique des systèmes flexionnels. L'objectif du présent article est précisément de procéder à une telle étude, et de montrer qu'elle permet de dégager des propriétés du système qui ne sont pas mises en avant par les méthodes plus traditionnelles.

### **III. Une approche instrumentée de la morphologie implicative**

Dans cette section, après avoir présenté à grands traits les caractéristiques de la conjugaison du portugais européen (§III.1), nous présentons sur la base d'exemples concrets la méthode qui a été employée pour identifier automatiquement des patrons d'alternance entre formes dans les paradigmes verbaux portugais et pour quantifier l'interprédictibilité entre les cases des paradigmes. On se donne comme point de départ un lexique de formes fléchies de grande taille, entièrement transcrit phonétiquement (§III.2). Pour chaque couple de cases du paradigme  $\langle \alpha, \beta \rangle$ , on identifie alors un ensemble de patrons d'alternance caractérisant la relation entre ces deux cases, et on quantifie la distribution de ces patrons dans le lexique (§III.3). Sur cette base, on évalue, pour toute forme remplissant la case  $\alpha$ , la probabilité des différentes formes correspondantes remplissant la case  $\beta$  (§III.4). La distribution de ces probabilités conditionnelles est résumée par l'entropie conditionnelle de la distribution, qui donne une évaluation globale de la difficulté à prédire le patron reliant  $\alpha$  à  $\beta$  à partir des propriétés phonologique de la forme remplissant la case  $\alpha$  (§III.5).

### III.1. La conjugaison du portugais européen

Le tableau 3 présente la forme générale d'un paradigme verbal du portugais. On distingue 5 tiroirs à l'indicatif (présent, imparfait, passé simple, plus-que-parfait, futur), un conditionnel, 3 tiroirs au subjonctif (passé, présent, futur). L'impératif est réputé avoir 5 personnes, mais les 3SG, 1PL et 3PL sont toujours syncrétiques avec les personnes correspondantes du subjonctif présent. L'inventaire des formes non-finies est plus difficile. En plus de l'infinitif ordinaire, il existe un infinitif personnel qui s'accorde en personne avec le sujet ; la 3SG est syncrétique avec l'infinitif ordinaire<sup>3</sup>. Certains verbes transitifs distinguent deux formes de participe passé, avec, selon la tradition, une spécialisation d'une des formes (la forme « régulière ») pour l'expression des temps composés, et de l'autre (la forme « irrégulière ») pour l'expression du passif périphrastique. Il semble cependant que dans la langue contemporaine ces deux formes soient en variation libre ; même dans le contexte formel de l'écriture journalistique, on trouve ainsi des exemples comme les suivants, tirés du corpus CETEMPúblico, où le participe régulier et le participe irrégulier se retrouvent tous deux aussi bien au passé composé (2b,2d) qu'au passif (2a,2c).

(2) a. Participe « irrégulier », emploi passif

A EDP era responsável já que era ela quem tinha **encarregue** os seus funcionários de responderem aos problemas que lhes eram apresentados pelos consumidores da electricidade [...]

L'EDP était responsable, dans la mesure où c'était elle qui avait chargé son personnel de répondre aux questions qui lui étaient soumises par les consommateurs d'électricité. (50246-soc-98a-2)

b. Participe « irrégulier », emploi actif

Em relação à Telepac, a Telecom foi **encarregue** de assumir a liderança de um processo de avaliação das várias empresas [...]

En ce qui concerne Telepac, Telecom était chargé d'assumer la direction d'une procédure d'évaluation de plusieurs sociétés. (215690-eco-93b-2)

c. Participe « régulier », emploi passif

[...] o juiz de instrução que foi **encarregado** de investigar o caso [...]

le juge d'instruction qui a été chargé d'enquêter sur l'affaire (13002-eco-96b-1)

d. Participe « régulier », emploi actif

[...] o primeiro-ministro António Guterres tinha **encarregado** os ministros da Agricultura e da Saúde de elaborar um plano de emergência conjunto [...]

Le Premier ministre António Guterres avait chargé les ministres de l'Agriculture et de la Santé de préparer un plan d'urgence commun. (628987-clt-96a-3)

---

<sup>3</sup> L'infinitif personnel est le plus souvent syncrétique avec le subjonctif futur, mais certains verbes les distinguent par une alternance radicale, par exemple ir « aller » : FUT.SBJV-2SG *for-es* vs. INF-2SG *ir-es*.

	1SG	2SG	3SG	1PL	2PL	3PL	
PRS.IND	util'izu	util'izeʃ	util'izə	utiliz'əmuʃ	utiliz'aiʃ	util'izẽũ	
PST.IPFV.IND	utiliz'avə	utiliz'avəʃ	utiliz'avə	utiliz'avəmuʃ	utiliz'avəiʃ	utiliz'avẽũ	
PST.PFV.IND	utiliz'ei	utiliz'ajtə	utiliz'o	utiliz'amuʃ	utiliz'ajtəʃ	utiliz'arẽũ	
PST.PRF.IND	utiliz'arə	utiliz'arəʃ	utiliz'arə	utiliz'arəmuʃ	utiliz'arəiʃ	utiliz'arẽũ	
FUT.IND	utilizər'ei	utilizər'aj	utilizər'a	utilizər'emuʃ	utilizər'eij	utilizər'ẽũ	
COND	utilizər'iə	utilizər'iəʃ	utilizər'iə	utilizər'iəmuʃ	utilizər'iəiʃ	utilizər'iẽũ	
PRS.SBJV	util'izə	util'izəʃ	util'izə	utiliz'emuʃ	utiliz'eij	util'izẽĩ	
PST.SBJV	utiliz'asə	utiliz'asəʃ	utiliz'asə	utiliz'asəmuʃ	utiliz'asəiʃ	utiliz'asẽĩ	
FUT.SBJV	utiliz'ar	utiliz'arəʃ	utiliz'ar	utiliz'armuʃ	utiliz'ardəʃ	utiliz'arẽĩ	
IMP	—	util'izə	util'izə	utiliz'emuʃ	utiliz'ai	util'izẽĩ	
PER.INF	utiliz'ar	utiliz'arəʃ	utiliz'ar	utiliz'armuʃ	utiliz'ardəʃ	utiliz'arẽĩ	
INF	GER	PST.PTCP	M.SG	M.PL	PASS.PTCP	F.SG	F.PL
utiliz'ar	utiliz'ẽdu	utiliz'adu	utiliz'adu	utiliz'adə	utiliz'aduʃ	utiliz'adəʃ	

**Tableau 3** La conjugaison du verbe UTILIZAR « utiliser »

La classification traditionnelle des verbes portugais les regroupe en 3 classes qui se distinguent par la voyelle thématique visible par exemple à l'infinitif, et qui peut être *-a*, *-e* ou *-i*. Celles-ci sont exemplifiées dans le tableau 4, qui montre les formes de l'indicatif présent de 3 verbes caractéristiques<sup>4</sup>. On identifie ici clairement 4 marqueurs de personne indépendants de la classe : 1SG /-u/, 2SG /-ʃ/, 1PL /-muʃ/, 2PL /-ʃ/<sup>5</sup>. La voyelle thématique est présente sous sa forme distinctive aux 1PL et 2PL, mais se manifeste sous une forme possiblement différente suivant les cas (/ə/ vs. /ai/, /e/ vs. /ei/). Aux 2SG et 3SG, la distinction entre les voyelles thématiques des deuxième et troisième conjugaisons est neutralisée. À la

<sup>4</sup> Pour les besoins de la présente discussion, on appelle *radical* la séquence qui précède l'éventuelle voyelle thématique, *marqueurs flexionnels* les morphes qui la suivent, et *terminaison* la sous-chaîne complète qui suit le radical, y compris l'éventuelle voyelle thématique.

<sup>5</sup> Pour fixer le vocabulaire descriptif, on admet la segmentation suivante pour les formes de présent 2PL :

- (i) Première conjugaison : X+ai+f
- (ii) Deuxième conjugaison : X+ei+f
- (iii) Troisième conjugaison : X+i+f

Cette décision, qui suit l'esprit de Boyé (2000), est motivée par le fait qu'elle permet d'établir un marqueur uniforme de la 2PL, sous la forme de /ʃ/, y compris pour les quelques verbes irréguliers qui ont une 2PL en /-dəʃ/, voir par exemple VIR dans le tableau 5. Cette segmentation ne va cependant pas de soi : une alternative plausible consisterait à poser des voyelles thématiques /a/ vs. /e/ vs. /i/, suivis d'un marqueur /-iʃ/ avec coalescence des deux /i/ à la 3<sup>e</sup> conjugaison. Dans cette segmentation alternative, la 2PL des verbes comme VIR devrait être considérée comme une forme fléchie supplétive (Boyé 2006).

La décision de segmentation est de peu de conséquence pour la présente analyse : une des vertus de l'approche basée sur les mots (et non sur les radicaux) utilisée ici est justement de ne pas nécessiter de choisir une segmentation uniforme. Voir sur ce point la section IV.4 ci-après.



3PL la voyelle thématique fusionne<sup>6</sup> avec le marqueur de personne et on obtient la réalisation d'une terminaison /-ẽũ/ pour la première conjugaison et de /-ẽĩ/ pour les deux autres. Enfin, la 1SG ne réalise pas de voyelle thématique.

La situation du présent est caractéristique de l'ensemble du paradigme, où l'on rencontre d'autres cas de variation du timbre de la voyelle thématique, de neutralisation partielle entre classes 2 et 3, ou de fusion. Dans l'ensemble on peut dès à présent observer que le système de voyelles thématiques contribue à améliorer la prédictibilité de la conjugaison, mais que les neutralisations, voire l'absence totale de voyelle thématique au PRS.1SG sont source d'opacité : si un locuteur n'a été exposé qu'à une forme de PRS.1SG d'un verbe, il ne peut inférer fiablement à quelle classe appartient ce verbe ; s'il est exposé à une 3SG en /-əʃ/, il sait que le verbe n'est pas de la première conjugaison, mais il pourrait être de la 2<sup>e</sup> ou de la 3<sup>e</sup>.

	1SG	2SG	3SG	1PL	2PL	3PL
UTILIZAR	util'izu	util'izəʃ	util'izə	utiliz'əmuʃ	utiliz'aiʃ	util'izẽũ
APRENDER	əpr'ẽdu	əpr'ẽdəʃ	əpr'ẽdə	əprẽd'emuʃ	əprẽd'eiʃ	əpr'ẽdẽĩ
IMPRIMIR	ĩpr'imu	ĩpr'iməʃ	ĩpr'imə	ĩprim'imuʃ	ĩprim'ij	ĩpr'imẽĩ

**Tableau 4** Le présent de l'indicatif de 3 verbes illustrant les 3 classes de conjugaison majeures

Les trois classes majeures de conjugaison n'épuisent pas la liste des patrons de conjugaison que l'on rencontre en portugais européen. Le tableau 5 donne quelques exemples supplémentaires de verbes au présent de l'indicatif qui mettent en jeu divers phénomènes notables : alternances vocaliques entre voyelles accentuées et inaccentuées (SECAR, REZAR), alternances de voyelles accentuées (DIVERTIR, DORMIR), allomorphie ou supplétion radicale (TRAZER, OUVIR, VIR), terminaison irrégulière (VIR, ESTAR). Encore une fois, le présent est représentatif de l'ensemble du paradigme, qui fournit d'autres exemples de phénomènes du même type.

	1SG	2SG	3SG	1PL	2PL	3PL
SECAR	s'eku	s'ekəʃ	s'ekə	sək'əmuʃ	sək'aiʃ	s'ekẽũ
REZAR	r'ezu	r'ezəʃ	r'ezə	rəz'əmuʃ	rəz'aiʃ	r'ezẽũ
DIVERTIR	div'irtu	div'ertəʃ	div'ertə	divərt'imuʃ	divərt'ij	div'ertẽĩ
DORMIR	d'urmu	d'orməʃ	d'ormə	durm'imuʃ	durm'ij	d'ormẽĩ
TRAZER	tr'agu	tr'azəʃ	tr'aj	trəz'emuʃ	trəz'eiʃ	tr'azẽĩ
OUVIR	'osu	'ovəʃ	'ovə	ov'imuʃ	ov'ij	'ovẽĩ
VIR	v'əɲu	v'ẽij	v'ẽĩ	v'imuʃ	v'ĩdəʃ	v'ẽĩẽĩ
ESTAR	əʃt'o	əʃt'aʃ	əʃt'a	əʃt'əmuʃ	əʃt'aiʃ	əʃt'ẽũ

**Tableau 5** Le présent de l'indicatif de quelques verbes illustrant d'autres patrons

### III.2. Le lexique utilisé

Beaucoup d'études des systèmes flexionnels s'appuient sur une première classification des données issue de la tradition grammaticale, et se contentent d'analyser un exemplaire de

<sup>6</sup> Mateus & d'Andrade (2000 :74-75) analysent les 3PL comme la combinaison d'une voyelle thématique avec un autosegment nasal. Quoi qu'on pense de l'analyse autosegmentale, il reste que la partie distinctive de la terminaison en est le dernier segment et non le premier ; il n'est donc pas possible de proposer une segmentation classique avec une voyelle thématique adjacente au radical.

chaque patron flexionnel identifié (voir par exemple Bonami & Boyé 2003, Ackerman & Malouf 2013, Stump & Finkel 2013). Cette approche, si elle a l'avantage de la facilité d'accès aux données, a trois inconvénients majeurs. D'abord, il est courant que la tradition rate certains contrastes subtils dans les données qui peuvent avoir des conséquences importantes pour l'analyse. Ensuite, l'analyse de lexèmes exemplaires ne permet pas de se constituer une image de l'ampleur des phénomènes étudiés, ce qui peut conduire à proposer des analyses inappropriées—par exemple reconfigurer la phonologie d'une langue pour rendre compte d'un phénomène qui affecte une poignée de mots. Enfin et surtout, elle interdit de produire une étude détaillée des propriétés phonotactiques caractérisant les lexèmes exemplifiant un patron flexionnel—ces propriétés pouvant conditionner les relations implicatives entre formes (Albright, 2002).

Ce travail est donc basé sur un lexique étendu, constitué de toutes les formes des 1000 verbes les plus fréquents dans le corpus CETEMPúblico (Santos & Rocha 2001). Ce lexique a été adapté à partir des données fournies par le dictionnaire de prononciation de l'Université de Coimbra (Veiga, A., Candeias, S. and Perdigão, 2012), en collaboration étroite avec deux de ses auteurs, Sara Candeias et Fernando Perdigão. Les transcriptions utilisées sont des transcriptions concrètes standardisées<sup>7</sup> : il existe bien sûr de la variabilité dans la prononciation des mots portugais européens, en fonction de la variété, du registre et du débit ; la variabilité affecte notamment la réalisation des voyelles inaccentuées. La transcription utilisée ici correspond à une réalisation possible en contexte formel avec un débit relativement lent, ce qui maximise la quantité de voyelles réalisées. Si les transcriptions ont en partie été construites de manière automatique en utilisant des technologies de synthèse vocale, une proportion importante des formes a été vérifiée à la main.

Une caractéristique importante du lexique que nous avons utilisé est qu'une seule forme est donnée par case du paradigme de chaque lexème : aucun phénomène de surabondance (Thornton, 2012) n'est pris en compte. Le lexique ignore donc en particulier les cas de participes passés multiples : les verbes concernés sont codés avec un participe unique, en l'occurrence le plus « régulier » selon la description traditionnelle. De même on n'a pas pris en compte la variation en genre et en nombre du participe passif, qui ne concerne que les verbes transitifs. Au total, on se retrouve donc avec des paradigmes de 68 formes, soit 68 000 formes fléchies.

### III.3 L'identification des patrons d'alternance

Nous nous attachons maintenant à la question d'étudier les relations implicatives qui mènent d'une case du paradigme à une autre case. Pour des lexèmes à 68 formes, il existe  $68 \times 67 = 4556$  couples de formes à étudier. Il n'est évidemment pas question de mener cette étude à la main (encore moins sur un lexique de 1000 verbes), et il est donc nécessaire de construire une méthode automatique. La méthode utilisée ici est présentée en détail dans Bonami & Boyé (sous presse) ; nous nous contenterons donc d'en résumer les caractéristiques les plus saillantes.

Le point de départ est d'inférer à partir d'un ensemble de couples de formes les patrons d'alternance qu'ils exhibent. Faute de disposer d'une méthode générique pour procéder à cette

---

<sup>7</sup> Les transcriptions adoptées sont proches de celles de Mateus & d'Andrade (2000), à trois différences près : (i) les semi-voyelles ne sont pas distinguées des voyelles hautes ; (ii) la voyelle centrale non-basse est transcrite [ə] plutôt que [i] ; (iii) l'accent est noté en utilisant le symbole de l'IPA placé immédiatement avant la voyelle accentuée.

identification<sup>8</sup>, nous nous appuyons de manière opportuniste sur le fait que la flexion du portugais est réputée être entièrement suffixante. Pour tout couple de formes, on considère donc qu'elles sont reliés par un patron qui consiste à substituer la plus courte sous-chaîne finale possible. Concrètement, le tableau 6 indique le résultat de cette opération pour les 20 lexèmes les plus fréquents du corpus, en ce qui concerne la relation du présent de l'indicatif 1SG au subjonctif présent 1SG. Le lecteur attentif notera que les patrons identifiés coïncident souvent, mais pas toujours, avec une segmentation entre radical et terminaison : pour les verbes manifestant de l'allomorphie radicale, la différence entre les deux radicaux est encodée dans le patron d'alternance. Cette situation n'est pas problématique, dans la mesure où l'enjeu de la présente procédure n'est pas de déterminer des frontières de morphes mais d'identifier des ensembles de lexèmes qui se fléchissent de la même manière : VER, par exemple, est correctement identifié comme unique en son genre (puisque'il possède une allomorphie radicale imprédictible), bien qu'il manifeste des terminaisons régulières.

lexème	IND.1SG	SBJV.1SG	patron	lexème	IND.1SG	SBJV.1SG	patron
SER	so	sɐizɐ	_o#~_ɐizɐ#	DEIXAR	d'ɛifʊ	d'ɛifə	_u#~_ə#
TER	tɛɾu	tɛɾɐ	_u#~_ɐ#	VIR	v'ɛɾu	v'ɛɾɐ	_u#~_ɐ#
FAZER	f'asu	f'asɐ	_u#~_ɐ#	ENCONTRAR	ẽk'õtru	ẽk'õtrɐ	_u#~_ə#
VER	v'ɛizu	v'ɛzɐ	_ɛizu#~_ɛzɐ#	HAYER <sup>9</sup>	'ɛi	'azɐ	#ɛi#~#azɐ#
ESTAR	ɛst'o	ɛst'ɛizɐ	_o#~_ɛizɐ#	SEGUIR	s'igu	s'igɐ	_u#~_ɐ#
PARTIR	p'artu	p'artɐ	_u#~_ɐ#	ENTRAR	'ẽtru	'ẽtrɐ	_u#~_ə#
PASSAR	p'asu	p'asɐ	_u#~_ə#	CONTINUAR	kõtin'uu	kõtin'uə	_u#~_ə#
CHEGAR	ʃ'egu	ʃ'egɐ	_u#~_ə#	MANTER	mẽt'ɛɾu	mẽt'ɛɾɐ	_u#~_ɐ#
FICAR	f'iku	f'ikɐ	_u#~_ə#	GANHAR	g'aju	g'ajɐ	_u#~_ə#
CRIAR	kr'iu	kr'ia	_u#~_ə#	TENTAR	t'ẽtu	t'ẽtɐ	_u#~_ə#

**Tableau 6** Patrons élémentaires reliant L'IND.PRS.1SG au SBJV.PRS.1SG pour 20 lexèmes

Le caractère contre-intuitif de la segmentation proposée est encore plus frappant si on examine des couples de formes où l'accent ne tombe pas sur la même syllabe, provoquant des réductions vocaliques dans l'une mais pas l'autre des deux formes concernées. Le tableau 7 illustre cette situation, toujours sur les mêmes exemples de verbes, en comparant le présent 1SG au présent 1PL. Dans ce cas la procédure actuelle identifie autant de patrons qu'il y a de combinaisons de voyelle alternante préthématique, de voyelle thématique et de séquence de consonnes intermédiaire ; la non identification des alternances vocaliques, qui sont un phénomène phonologique régulier de la langue, conduit donc à introduire 5 patrons supplémentaires qui pourraient être réduits par une analyse plus fine<sup>10</sup> ; la colonne « patron inférable » indique le patron plus abstrait qu'on voudrait voir identifié, mais que la procédure actuellement utilisée n'identifie pas<sup>11</sup>.

<sup>8</sup> La difficulté majeure pour ce faire est de disposer d'un algorithme qui identifie correctement des patrons exhibant différents types d'alternances morphologiques ; par exemple, qui, face à des données de l'arabe classique, infère correctement que (*kalaba*, *kuliba*) doit être analysé comme une transfixation, plutôt qu'une infixation de *-ala-* vs. *-uli-*, une suffixation de *-alaba* vs. *-uliba*, ou une préfixation de *kala-* vs. *kuli-*. L'utilisation d'un modèle explicite de l'analogie formelle (voir p. ex. Lavallée et Langlais 2011) est une piste prometteuse mais qui demande encore à être testée à grande échelle.

<sup>9</sup> En dehors de la 3SG, HAYER est toujours suivi de la forme *de* /dɐ/.

<sup>10</sup> Sur l'ensemble du lexique étudié on introduit de la sorte 122 patrons apparents qui sont tous réductibles aux 3 patrons les plus fréquents par prise en compte des alternances vocaliques régulières.

<sup>11</sup> La non-identification des patrons abstraits a des conséquences sur certains des résultats quantitatifs présentés ci-après dans la section IV. Voir la note 20.

lexème	PRS.1SG	PRS.1PL	patron identifié	patron inférable
SER	so	somuʃ	_#~_muʃ#	
TER	tɛʁu	temuʃ	_ɛʁu #~_emuʃ#	
FAZER	f'asu	fɛz'emuʃ	_asu#~_ɛz'emuʃ#	
VER	v'ɛizʊ	v'emuʃ	_ɛizʊ#~_emuʃ#	
ESTAR	ɛʃt'o	ɛʃt'emuʃ	_o#~_ɛmu #	
PARTIR	p'artu	pɛrt'imuʃ	_artu #~_ɛrt'imuʃ#	_u#~_imuʃ#
PASSAR	p'asu	pɛs'emuʃ	_asu#~_ɛs'emuʃ#	_u#~_ɛmuʃ#
CHEGAR	ʃ'egu	ʃɔg'emuʃ	_egu #~_ɔg'emuʃ#	_u#~_ɛmuʃ#
FICAR	f'iku	fik'emuʃ	_u#~_ɛmuʃ#	
CRIAR	kr'iu	kri'emuʃ	_u#~_ɛmuʃ#	
DEIXAR	d'ɛiʃu	dɛiʃ'emuʃ	_u#~_ɛmuʃ#	
VIR	v'ɛʁu	v'imuʃ	_ɛʁu #~_imuʃ#	
ENCONTRAR	ɛk'ɔtru	ɛkɔtr'emuʃ	_u#~_ɛmuʃ#	
HAYER	'ɛi	ɛv'emuʃ	_i#~_vemuʃ#	
SEGUIR	s'igu	sɔg'imuʃ	_u#~_imuʃ#	
ENTRAR	'ɛtru	ɛtr'emuʃ	_u#~_ɛmuʃ#	
CONTINUAR	kɔtin'uu	kɔtinu'emuʃ	_u#~_ɛmuʃ#	
MANTER	mɛt'ɛʁu	mɛt'emuʃ	_ɛʁu#~_emuʃ#	
GANHAR	g'aju	gaj'emuʃ	_u#~_ɛmuʃ#	
TENTAR	t'ɛtu	tɛt'emuʃ	_u#~_ɛmuʃ#	

**Tableau 7** Patrons élémentaires reliant L'IND.PRS.1SG à l'IND.PRS.1PL pour 20 lexèmes

À la suite de cette première phase, on procède à une classification des lexèmes en fonction des patrons qu'ils instancient. En plus de compter les lexèmes, on procède à une description des caractéristiques phonotactiques que les membres d'une même classe partagent. Cette description est construite en utilisant la stratégie de généralisation minimale définie par Albright (2002) : en partant de la frontière de la partie affectée par le changement, on construit segment par segment la description en termes de traits distinctifs la plus spécifique qui vaille pour l'ensemble des lexèmes concernés. Dans le cas de la relation entre présent et subjonctif 1SG, on aboutit pour les 20 verbes du tableau 6 aux résultats indiqués dans le tableau 8. Dans le cas des verbes réguliers de la première conjugaison la diversité phonotactique est trop grande pour qu'on puisse identifier des généralisations. Pour les verbes des 2<sup>e</sup> et 3<sup>e</sup> conjugaisons (les deux conjugaisons ne se différencient pas pour les deux formes qui nous intéressent) on peut cependant identifier une généralité sur les 6 exemples présents : tous se terminent par une séquence constituée d'une consonne ou d'une voyelle non-haute, d'une sonorante non-nasale non-latérale, et d'une consonne. Si on ne se limite plus aux 20 premiers verbes mais qu'on regarde l'ensemble des données (tableau 9) cependant, il s'avère que cette généralisation ne tient pas, et que le patron avec subjonctif en /ɛ/ n'impose pas de condition phonotactique sur le contexte. On constate ensuite une généralisation intéressante concernant les verbes à IND.PRS.1SG en /-o/ : les deux verbes concernés (SER et ESTAR) terminent leur radical par une obstruante sourde dentale. Si on regarde l'ensemble des verbes (tableau 9), on constate qu'il n'y a pas d'autres verbes à suffixe /-o/, et que donc cette observation est générale. Enfin, on identifie quelques patrons qui ne s'appliquent qu'à un seul lexème ; la description du contexte phonotactique est maximale spécifiquement, puisqu'aucune généralisation n'a pu être faite. Un des trois patrons concernés s'avère avoir sur le lexique complet quelques autres représentants (il s'agit de VER et de ses dérivés).

patron	effectif
u → ə / X ___ #	10
u → ɐ / X [-nas,-haut] [+son,-nas,-lat] [+cons] ___ #	6
o → ɛiʒɐ / [+cons,-voix,+cor,+ant] ___ #	2
ɛiʒu → ɛʒɐ / #v ___ #	1
ɛi → aʒɐ / # ___ #	1

**Tableau 8** Patrons d’alternance du PRS.IND.1SG au SBJV.IND.1SG, avec généralisation minimale sur les propriétés phonotactiques du contexte (20 verbes les plus fréquents)

patron	effectif
u → ə / X ___ #	703
u → ɐ / X ___ #	289
o → ɛiʒɐ / [+cons,-voix,+cor,+ant] ___ #	2
ɛiʒu → ɛʒɐ / Xv ___ #	3
ɛɾu → ɛiɾɐ / Xk ___ #	2
ɛi → aʒɐ / # ___ #	1

**Tableau 9** Patrons d’alternance du PRS.IND.1SG au SBJV.IND.1SG, avec généralisation minimale sur les propriétés phonotactiques du contexte (1000 verbes les plus fréquents)

La procédure d’indentification des patrons d’alternance décrite dans cette section est perfectible par bien des aspects, que nous avons tenté d’expliquer. Elle a cependant deux avantages notables. D’une part, elle est très simple, et donc les résultats sont très faciles à examiner, analyser, et critiquer. D’autre part, elle peut être implémentée de manière très efficace, et donc appliquée sans difficultés à de gros ensembles de données.

### III.4 Mesurer la prédictibilité

Quelle que soit la méthode de classification des patrons d’alternance entre formes utilisée, on peut en déduire des informations sur la prédictibilité d’une forme à partir d’une autre. Considérons de nouveau les informations comprises dans le tableau 9. Il y apparaît clairement plusieurs généralisations. D’abord, un seul des patrons identifiés est compatible avec le fait que l’IND.PRS.1SG se termine en /-o/. Face à une forme en /-o/, on peut donc prédire de manière catégorique que le SBJV.PRS.1SG sera en /-ɛiʒɐ/. Ensuite, si un verbe se termine en /-u/ mais ni en /-ɛiʒu/ ni en /-ɛɾu/ à l’IND.PRS.1SG, deux terminaisons sont possibles au SBJV.PRS.1SG : /-ə/ ou /-ɐ/. Que peut-on en déduire? On ne peut pas prédire de manière catégorique la terminaison du subjonctif sur la base de l’indicatif, mais on prédit qu’il n’y a que deux possibilités, et la fréquence des deux patrons nous donne une indication sur la probabilité de chaque alternative: face à un verbe inconnu à IND.PRS.1SG en /-u/, la probabilité que son SBJV.PRS.1SG soit en /-ə/ est proche de  $703/(703+289) \approx 71\%$ .

On voit donc que la connaissance des patrons d’alternance et de leur distribution nous donne des informations cruciales sur la prédictibilité d’une forme à partir d’une autre. Il y a cependant des informations pertinentes qui ne sont pas contenues dans le tableau 9. Prenons le cas des verbes qui se terminent en /-ɛɾu/ à l’IND.PRS.1SG. À en juger par le tableau 9, trois

possibilités sont ouvertes en ce qui concerne leur SBJV.PRS.1SG : /-εrə/ (application du premier patron), /-εrɐ/ (application du second patron), ou /-εirɐ/ (application du cinquième patron). Le tableau ne nous donne cependant pas les informations pertinentes pour évaluer la probabilité relative de chaque possibilité: on sait combien de verbes en /-εru/ exemplifient le cinquième patron, mais pas les autres.

Pour accéder à cette information, on doit maintenant procéder à une classification des lexèmes non pas en fonction du patron d'alternance qu'ils instancient, mais en fonction des patrons qu'ils *pourraient* instancier étant donné l'aspect phonologique de leur IND.PRS.1SG. On doit faire un nouvel examen complet du lexique, et se demander par exemple quels verbes sont susceptibles d'instancier le patron [εru → εirɐ / Xk \_\_\_ #], autrement dit quels verbes se terminent en /-kεru/ à l'IND.PRS.1SG. Il s'avère que s'il existe 16 verbes en /-εru/, seuls 2 sont en /-kεru/, à savoir justement les deux verbes QUERER et REQUERER.

Le tableau 10 donne les détails de cette exploration pour l'ensemble des 1000 lexèmes étudiés. Les lexèmes sont groupés en classes de similarité phonologique à l'IND.PRS.1SG, en fonction de l'inventaire des patrons qu'ils pourraient théoriquement instancier. L'effectif de chaque classe est noté, ainsi que le nombre de lexèmes instanciant chacun des patrons envisageables. Il est important de retenir que, si la classification est entièrement basée sur la forme phonologique, les classes ne sont pas nécessairement phonologiquement naturelles : par exemple la classe 1 est la classe des verbes dont l'IND.PRS.1SG se termine en /u/ mais ni en /εizu/ ni en /εru/. Cette non-naturalité phonologique des classes est revendiquée : les classes sont déterminées automatiquement en fonction de l'inventaire des patrons d'alternance rencontrés, et non en fonction de préconceptions sur la sensibilité des alternances à telle ou telle propriété phonologique.

classe	exemple	effectif	patron	exemple	effectif
c <sub>1</sub>	UTILIZAR,	992	u → ə / X ___ #	util'izu ~ util'izə	703
	DEDUZIR		u → ɐ / X ___ #	dəd'uzu ~ dəd'uzɐ	289
c <sub>2</sub>	PREVER	3	εizu → eʒɐ / Xv ___ #	prəv'εizu ~ prəv'eʒɐ	3
			u → ə / X ___ #	—	0
			u → ɐ / X ___ #	—	0
c <sub>3</sub>	REQUERER	2	εru → εirɐ / Xk ___ #	rək'εru ~ rək'εirɐ	2
			u → ə / X ___ #	—	0
			u → ɐ / X ___ #	—	0
c <sub>4</sub>	ESTAR	2	o → εizɐ / #... ___ #	əʃt'o ~ əʃt'εizɐ	2
c <sub>5</sub>	HAYER	1	εi → aʒɐ / # ___ #	'εi ~ 'aʒɐ	1

**Tableau 10** Distribution des patrons d'alternance envisageables par type de finale

La méthode étant globalement établie, examinons les résultats qu'elle donne sur quelques exemples caractéristiques. D'abord, il est important de noter que les relations implicatives entre deux formes sont généralement orientées : ce qui est prédictible dans une direction peut être imprédictible dans l'autre, et réciproquement. L'examen des relations implicatives du SBJV.PRS.1SG vers l'IND.PRS.1SG illustre clairement ce point. Comme l'illustre le tableau 11, toute l'incertitude présente dans la direction opposée est ici absente : il n'est pas possible de prédire la voyelle thématique du SBJV.PRS.1SG à partir de la forme athématique de l'IND.PRS.1SG, mais à l'inverse la suppression de la voyelle thématique du SBJV.PRS.1SG à

l'IND.PRS.1SG donne un résultat catégoriquement prédictible. Il s'en suit que l'IND.PRS.1SG est toujours catégoriquement prédictible à partir du SBJV.PRS.1SG.

classe	exemple	effectif	patron	exemple	effectif
$c_1$	UTILIZAR	705	$\text{ə} \rightarrow \text{u} / \text{X} \_ \#$	util'izə ~ util'izu	705
$c_2$	DEDUZIR	288	$\text{ɐ} \rightarrow \text{u} / \text{X} \_ \#$	dəd'uzɐ ~ dəd'uzu	288
$c_3$	PREVER	3	$\text{ɛzɐ} \rightarrow \text{ɛizu} / \text{Xv} \_ \#$	prəv'ɛzɐ ~ prəv'ɛizu	3
			$\text{ɐ} \rightarrow \text{u} / \text{X} \_ \#$	—	0
$c_4$	REQUERER	2	$\text{ɛirɐ} \rightarrow \text{ɛru} / \text{Xk} \_ \#$	rək'ɛirɐ ~ rək'ɛru	2
			$\text{ɐ} \rightarrow \text{u} / \text{X} \_ \#$	—	0
$c_5$	ESTAR	2	$\text{ɛizɐ} \rightarrow \text{o} / \# \dots \_ \#$	ɛft'ɛizɐ ~ ɛft'o	2
			$\text{ɐ} \rightarrow \text{u} / \text{X} \_ \#$	—	0
$c_6$	HAVER	1	$\text{aɜɐ} \rightarrow \text{ɛi} / \# \_ \#$	'aɜɐ ~ 'ɛi	1
			$\text{ɐ} \rightarrow \text{u} / \text{X} \_ \#$	—	0

**Tableau 11** Distribution des patrons d'alternance pour [SUBJ.PRS.1SG  $\Rightarrow$  IND.PRS.1SG]

Dans certains cas limites, l'interprédicibilité est parfaite dans les deux directions entre deux cases du paradigme. Un exemple simple d'une telle situation se rencontre entre l'infinitif et l'IND.PRS.1PL. Les tableaux 12 et 13 indiquent les patrons d'alternance et leur distribution dans les deux directions. On voit que dans les deux cas, aucune incertitude ne peut intervenir : les deux premiers patrons ont des contextes d'application mutuellement exclusifs, et le troisième n'est applicable qu'à un verbe.

classe	exemple	effectif	patron	exemple	effectif
$c_1$	UTILIZAR	705	$\text{ar} \rightarrow \text{ɐmuɣ} / \text{X} \_ \#$	utiliz'ar ~ utiliz'ɐmuɣ	704
$c_2$	DEDUZIR	288	$\text{r} \rightarrow \text{muɣ} / \text{X} [\text{+voc}, -\text{bas}] \_ \#$	dəd'uzir ~ dəd'uzimuɣ	295
			$\text{er} \rightarrow \text{omuɣ} / \#s \_ \#$	s'er ~ s'omuɣ	1
$c_3$	SER	1	$\text{r} \rightarrow \text{muɣ} / \text{X} [\text{+voc}, -\text{bas}] \_ \#$	—	0

**Tableau 12** Distribution des patrons d'alternance pour [INF  $\Rightarrow$  IND.PRS.1PL]

classe	exemple	effectif	patron	exemple	effectif
$c_1$	UTILIZAR	705	$\text{ɐmuɣ} \rightarrow \text{ar} / \text{X} \_ \#$	utiliz'ɐmuɣ ~ utiliz'ar	704
$c_2$	DEDUZIR	288	$\text{muɣ} \rightarrow \text{r} / \text{X} [\text{+voc}, -\text{bas}] \_ \#$	dəd'uzimuɣ ~ dəd'uzir	295
			$\text{omuɣ} \rightarrow \text{er} / \#s \_ \#$	s'omuɣ ~ s'er	1
$c_3$	SER	1	$\text{muɣ} \rightarrow \text{r} / \text{X} [\text{+voc}, -\text{bas}] \_ \#$	—	0

**Tableau 13** Distribution des patrons d'alternance pour [IND.PRS.1PL  $\Rightarrow$  INF]

Les exemples ci-dessus ont été choisis pour leur simplicité : dans le cas général, entre deux cases du paradigme  $X$  et  $Y$ , il existe un grand nombre de patrons d'alternance dans

chacune des deux directions, et un degré élevé d'imprédictibilité. De ce fait, l'examen individuel des tables de distribution des patrons est une tâche longue et délicate, qu'on souhaite ne devoir entreprendre que là où c'est réellement utile.

### III.5 L'entropie conditionnelle comme mesure globale de la prédictibilité

Les tableaux de distribution des patrons d'alternance présentés ci-dessus donnent une image assez détaillée de la structure implicative d'un petit recoin de la conjugaison du portugais européen. Quand on veut étudier le système dans son ensemble cependant, l'examen manuel de ces tableaux n'est pas une solution viable : avec un paradigme à 68 formes, ce sont  $68 \times 67 = 4556$  tableaux qu'il faudrait examiner. Il est donc indispensable de disposer de mesures quantitatives résumant l'information présente dans ces tableaux. À la suite d'Ackerman, Blevins & Malouf (2009), nous utilisons l'*entropie conditionnelle* comme mesure quantitative donnant une image globale de la prédictibilité d'une case d'un paradigme à partir d'une autre.

La motivation du choix de cette mesure va au delà de ce qui peut être développé dans cet article. Nous nous contenterons donc de donner quelques éléments d'explication qui permettront au lecteur de comprendre comment l'entropie conditionnelle est calculée et renvoyons à Ackerman et al. (2009) pour plus de détails.

L'idée de départ est de donner une interprétation explicitement probabiliste aux distributions de patrons d'alternance étudiées ci-dessus. Considérons d'abord les chiffres du tableau 9. En supposant que l'échantillon de verbes étudiés est représentatif de l'ensemble, ce tableau nous permet d'estimer la probabilité, pour un verbe tiré au hasard, d'instancier chacun des patrons de conjugaison identifiés. Les probabilités ainsi calculées sont données en (3).

- (3) a.  $P([\text{IND.PRS.1SG} \Rightarrow \text{SBJV.PRS.1SG}] : [\text{u} \rightarrow \text{ə}]) \approx 703/1000 = 0,703$   
 b.  $P([\text{IND.PRS.1SG} \Rightarrow \text{SBJV.PRS.1SG}] : [\text{u} \rightarrow \text{ɐ}]) \approx 289/1000 = 0,289$   
 c.  $P([\text{IND.PRS.1SG} \Rightarrow \text{SBJV.PRS.1SG}] : [\text{ɛizʊ} \rightarrow \text{ɛzɐ}]) \approx 3/1000 = 0,003$   
 d.  $P([\text{IND.PRS.1SG} \Rightarrow \text{SBJV.PRS.1SG}] : [\text{ɛru} \rightarrow \text{ɛirɐ}]) \approx 2/1000 = 0,002$   
 e.  $P([\text{IND.PRS.1SG} \Rightarrow \text{SBJV.PRS.1SG}] : [\text{o} \rightarrow \text{ɛizɐ}]) \approx 2/1000 = 0,002$   
 f.  $P([\text{IND.PRS.1SG} \Rightarrow \text{SBJV.PRS.1SG}] : [\text{ɛi} \rightarrow \text{azɐ}]) \approx 1/1000 = 0,001$

Considérons ensuite les effectifs des classes de formes d'IND.PRS.1SG apparaissant dans la partie gauche du tableau 10. Ceux-ci nous permettent d'estimer la probabilité, si on tire un lexème au hasard, que son IND.PRS.1SG appartienne à l'une ou l'autre des classes concernées.

- (4) a.  $P(\text{IND.PRS.1SG} \in c_1) \approx 992/1000 = 0,992$   
 b.  $P(\text{IND.PRS.1SG} \in c_2) \approx 3/1000 = 0,003$   
 c.  $P(\text{IND.PRS.1SG} \in c_3) \approx 2/1000 = 0,002$   
 d.  $P(\text{IND.PRS.1SG} \in c_4) \approx 2/1000 = 0,002$   
 e.  $P(\text{IND.PRS.1SG} \in c_5) \approx 1/1000 = 0,001$

On peut enfin interpréter en termes probabilistes les chiffres appartenant dans la dernière colonne du tableau 10. Ceux-ci correspondent à des probabilités conditionnelles: par exemple, la première ligne note le fait que si on se restreint à la classe  $c_1$ ,  $703/992 \approx 70,9\%$  des verbes instancient le patron  $[\text{u} \rightarrow \text{ə}]$ . On peut en inférer que la probabilité conditionnelle d'instancier le patron  $[\text{u} \rightarrow \text{ə}]$  sachant qu'on appartient à la classe  $c_1$  est proche de 0,709.



Le tableau 13 indique les probabilités conditionnelles ainsi déductibles du tableau 10. On notera qu'une probabilité conditionnelle peut être nulle pour deux raisons : soit parce que le patron, bien qu'applicable en principe à cette classe, se trouve n'être instancié par aucun membre de la classe<sup>12</sup> ; soit parce que le patron est simplement inapplicable. Les cases correspondant à cette dernière situation sont grisées dans le tableau.

	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$
$u \rightarrow \emptyset / X \_ \#$	0,709	0	0	0	0
$u \rightarrow \text{v} / X \_ \#$	0,281	0	0	0	0
$\text{v}i\text{z}u \rightarrow \text{e}z\text{v} / X\text{v} \_ \#$	0	1	0	0	0
$\text{e}r\text{u} \rightarrow \text{v}i\text{r}\text{e} \text{e}z\text{v} / X\text{k} \_ \#$	0	0	1	0	0
$o \rightarrow \text{v}i\text{z}\text{v} / [+cons, -voix, +cor, +ant] \_ \#$	0	0	0	1	0
$\text{v}i \rightarrow \text{a}z\text{v} / \# \_ \#$	0	0	0	0	1

**Tableau 14** Probabilité conditionnelle d'un patron [IND.PRS.1SG  $\Rightarrow$  SBJV.PRS.1SG] sachant la classe de IND.PRS.1SG

Le tableau 14 donne une interprétation probabiliste explicite aux observations sur la prédictibilité que nous avons déduite du tableau 9. Cette interprétation a l'avantage de permettre d'utiliser les outils de la théorie de l'information pour synthétiser la prédictibilité de la relation entre deux cases du paradigme. Etant donnée une variable aléatoire quelconque  $X$ , l'ENTROPIE de cette variable aléatoire, notée  $H(X)$  est définie comme en (5).

$$(5) \quad H(X) = - \sum_{x \in X} P(x) \log_2 P(x)$$

Intuitivement, l'entropie capte l'incertitude liée au choix d'une valeur pour une variable aléatoire. Imaginons une situation où la variable a deux valeurs possibles également probables. Dans ce cas, comme le montre le calcul en (6), l'entropie est de 1.

$$(6) \quad H(X) = - \left( \frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) = - \log_2 \frac{1}{2} = \log_2 2 = 1$$

Si l'une des valeurs possibles est plus probable que l'autre, l'entropie décroît : comme on le voit en (7), si on a deux valeurs possibles  $x_1$  et  $x_2$  mais que  $x_1$  est trois fois plus probable que  $x_2$ , l'entropie descend en dessous de 1.

(7)

$$H(X) = - \left( \frac{3}{4} \log_2 \frac{3}{4} + \frac{1}{4} \log_2 \frac{1}{4} \right) = - \left( \frac{3}{4} \log_2 3 - \frac{3}{4} \log_2 4 - \frac{1}{4} \log_2 4 \right) = 2 - \frac{3}{4} \log_2 3 \approx 0,8113$$

<sup>12</sup> Aux cellules claires correspond un biais possible dans l'estimation des probabilités : il se trouve que l'échantillon de verbes examiné ne contient aucun verbe à IND.PRS.1SG en /-vɛiʒu/ qui instancie un des deux patrons réguliers, mais l'examen d'un échantillon plus large pourrait montrer que cela n'est pas général. Les cellules grisées ne sont pas susceptibles de ce type de biais.

Dans le cas extrême illustré en (8) où l'une des deux valeurs, bien que théoriquement concevable, n'est jamais attestée en pratique, l'entropie tombe à  $0^{13}$  : une entropie nulle correspond donc à une situation de certitude.

$$(8) \quad H(X) = -(0 \times \log_2 0 + \log_2 1) = -\log_2 1 = 0$$

Si le nombre de valeurs possibles pour la variable aléatoire croît, toutes choses égales par ailleurs, l'entropie croît également : par exemple, s'il y a 4 valeurs possibles également probables, comme illustré en (9), l'entropie est de 2.

$$(9) \quad H(X) = -\left(\frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4}\right) = -\log_2 \frac{1}{4} = \log_2 4 = 2$$

Pour prendre un exemple concret, revenons sur les chiffres en (3), et demandons-nous quelle est l'incertitude liée à la détermination de la classe d'un IND.PRS.1SG tiré au hasard dans notre liste de 1000 verbes. Il y a 5 possibilités (chacune des classes  $c_1, c_2, c_3, c_4$  et  $c_5$ ). Le calcul en (10) nous donne une entropie très proche de 0 dans ce cas, ce qui correspond à l'intuition que si on tire un verbe au hasard, il y a très peu d'incertitude quant à la terminaison de son IND.PRS.1SG.

$$(10) \quad \begin{aligned} H(X) &= -(0,992 \times \log_2 0,992 + 0,003 \log_2 0,003 + 0,004 \log_2 0,002 + 0,001 \log_2 0,001) \\ &\approx 0.0825 \end{aligned}$$

De la notion d'entropie se déduit la notion d'entropie conditionnelle. Etant données deux variables aléatoires  $X$  et  $Y$ , l'entropie conditionnelle de  $Y$  sachant  $X$ , notée  $H(Y|X)$ , se définit comme en (11).

$$(11) \quad H(Y|X) = -\sum_{x \in X} P(x) \sum_{y \in Y} P(y|x) \log_2 P(y|x)$$

Intuitivement, l'entropie conditionnelle capte la dépendance entre l'incertitude associée à deux variables aléatoires : si la valeur de  $X$  conditionne fortement la valeur de  $Y$ , l'entropie conditionnelle sera basse ; si la valeur de  $X$  conditionne peu la valeur de  $Y$ , l'entropie sera plus haute. Le calcul prend la forme d'une somme pondérée des entropies calculées pour chaque valeur possible de la variable  $X$ . Dans le cas limite où les valeurs de  $X$  et de  $Y$  varient de manière totalement indépendante  $H(Y|X) = H(Y)$  ; à l'inverse, si la valeur de  $X$  détermine totalement la valeur de  $Y$ ,  $H(Y|X) = 0$ .

Nous sommes maintenant en position d'utiliser l'entropie pour fournir une mesure globale de la prédictibilité au sein des paradigmes. Spécifiquement, on peut calculer l'entropie conditionnelle du choix du patron d'alternance étant donné le choix de la classe de la forme de départ. Par exemple, à partir de (4) et du tableau 14, on calcule l'entropie conditionnelle du patron mis en jeu par la relation [IND.PRS.1SG  $\Rightarrow$  SBJV.PRS.1SG] étant donnée la classe de l'IND.PRS.1SG :

---

<sup>13</sup> Par convention, on considère dans les calculs d'entropie que  $0 \times \log 0 = 0$

$$(12) \quad H([\text{IND.PRS.1SG} \Rightarrow \text{SBJV.PRS.1SG}]|\text{IND.PRS.1SG}) =$$

$$\left( \begin{array}{l} 0,992(0,709 \log_2 0,709 + 0,281 \log_2 0,281) + \\ 0,003(\log_2 1) + \\ - 0,002(\log_2 1) + \\ 0,002(\log_2 1) + \\ 0,001(\log_2 1) \end{array} \right) =$$

$$-0,992(0,709 \log_2 0,709 + 0,281 \log_2 0,281) \approx 0,8537$$

Le même calcul appliqué aux données du tableau 11 fait aboutir à une valeur de zéro pour  $H([\text{SBJV.PRS.1SG} \Rightarrow \text{IND.PRS.1SG}]|\text{SBJV.PRS.1SG})$  : chaque classe ne contenant qu'un seul patron à probabilité non nulle, chacun des termes de la somme a pour valeur zéro. On capte donc bien la différence entre une prédictibilité totale et une prédictibilité partielle, et on est en mesure de différencier quantitativement des degrés de prédictibilité partielle.

Il est important de ne pas perdre de vue que l'entropie fournit un résumé d'une distribution de probabilités conditionnelles, et que, comme tout résumé, il fournit moins d'information que ce qu'il résume. Examiner les valeurs d'entropie conditionnelle reliant les cases d'un paradigme est donc toujours un raccourci, et il faut aller examiner les tables de probabilité pour comprendre les raisons linguistiques qui amènent à un certain chiffre. Ce sera la démarche de la section suivante : à partir d'une vision macroscopique des paradigmes du portugais européen induite par l'examen de la distribution des valeurs d'entropie conditionnelle, nous examinerons plus en détail certains cas caractéristiques pour déterminer les causes d'une prédictibilité particulièrement haute ou particulièrement basse.

#### IV. Premiers résultats

Dans cette section nous présentons une première interprétation des résultats produits par l'analyse automatique des relations implicatives entre les 4556 couples de cases des paradigmes du portugais européen. Il n'est bien sûr pas possible d'examiner tous les cas un par un, et nous allons donc sélectionner des informations à examiner.

##### IV.1 La prédictibilité totale

Commençons par examiner les exemples d'implications catégoriques. Une première observation importante est que sur les 4556 couples de cases, l'entropie est nulle dans 1174 cas, soit environ dans 26% des cas. Autrement dit, si on tire au hasard deux cases du paradigme, dans plus d'un cas sur quatre la forme de la seconde case est prédictible sur la base de celle de la première. Si maintenant on s'intéresse aux paires de cases entre lesquelles il y a interprédictibilité mutuelle<sup>14</sup>, il y en a 399 sur 2278, soit environ 18% des cas. On est donc fort loin d'une situation où la connaissance d'une forme d'un verbe permettrait à tout coup de prédire ses autres formes.

---

<sup>14</sup> Rappelons qu'une *paire* est un ensemble de deux éléments, alors qu'un *couple* est une séquence : il n'y a pas d'ordre dans une paire, il y a un ordre entre les éléments d'un couple. À partir d'un ensemble d'étiquettes de cases de paradigmes, on obtient donc toujours deux fois plus de couples que de paires.

Si on examine les situations qui favorisent l'interprédictibilité totale, il apparaît clairement que celle-ci est le plus souvent due à une situation qui rappelle la morphologie flexionnelle canonique (Corbett, 2007) :

- (13) Conditions suffisantes pour une relation d'interprédictibilité parfaite
- Les deux cases du paradigme ont l'accent dans la même position.
  - Il y a une démarcation régulière sur l'ensemble du lexique entre une partie constante (un pseudo-radical) et une partie variable (une pseudo-terminaison)<sup>15</sup>.
  - La partie radicale est identique pour les deux formes.
  - Les voyelles thématiques, là où elles existent, sont en co-variation d'une case à l'autre.

L'examen des formes de l'imparfait (Tableau 15) illustre ce point : l'accent est à toutes les formes sur la voyelle thématique. Cette voyelle thématique n'est pas constante, mais connaît pour chaque case deux variantes qui sont en co-variation : les lexèmes qui ont un /a/ au 1SG gardent ce /a/ dans tout le tiroir, ceux qui ont un /iə/ le gardent dans tout le tiroir sauf au 3PL où ils ont un /iẽ/. Enfin, il n'y a aucune allomorphie radicale.

	1SG	2SG	3SG	1PL	2PL	3PL
UTILIZAR	utiliz'avə	utiliz'avəf	utiliz'avə	utiliz'avəmuɸ	utiliz'avəɸ	utiliz'avẽũ
APRENDER	əprəd'ie	əprəd'ieɸ	əprəd'ie	əprəd'iemuɸ	əprəd'ieɸ	əprəd'iẽũ
IMPRIMIR	ĩprim'ie	ĩprim'ieɸ	ĩprim'ie	ĩprim'iemuɸ	ĩprim'ieɸ	ĩprim'iẽũ

**Tableau 15** Exemples représentatifs de la conjugaison de l'imparfait

Tous les cas de prédictibilité totale n'adhèrent pas aux conditions (13). Une exception particulièrement éclairante est rencontrée au prétérit perfectif (Tableau 16). Les verbes réguliers ont à ce tiroir 5 formes directement analogues, avec accent sur la voyelle thématique et contraste entre trois qualités de cette voyelle. La 1SG fait exception, en ce qu'elle neutralise la distinction entre les seconde et troisième conjugaisons ; elle n'est donc pas un bon prédicteur du reste du sous-paradigme. Si on examine les verbes irréguliers, on constate cependant, aussi bien à la 1SG qu'à la 3SG, un certain nombre de verbes qui font exception à (13) : la terminaison et/ou le radical attendus ne sont pas utilisés. Il est intéressant de noter que ces exemples ne sont pas de nature à faire baisser la prédictibilité, à cause de leurs propriétés phonotactiques distinctives. Toutes les formes incriminées ont une propriété qui permet de les distinguer du cas régulier. Au 1SG, la terminaison est une consonne ou /i/ ; au 3SG, c'est une consonne ou /ə/ ; dans les quatre cases restantes, la voyelle thématique est /ɛ/.

<sup>15</sup> Nous parlons de pseudo-radical et de pseudo-terminaison parce que la segmentation en question peut ne faire sens que pour les deux cases en question du paradigme, et ne pas s'étendre au paradigme dans son ensemble, même dans des cas où descriptivement on ne voudrait pas dire qu'il y a allomorphie radicale.

	1SG	2SG	3SG	1PL	2PL	3PL
UTILIZAR	utiliz'ei	utiliz'əftə	utiliz'o	utiliz'əmuɟ	utiliz'əftəɟ	utiliz'arɛũ
APRENDER	ɛprɛd'i	ɛprɛd'ɛftə	ɛprɛd'eu	ɛprɛd'əmuɟ	ɛprɛd'ɛftəɟ	ɛprɛd'erɛũ
IMPRIMIR	ĩprim'i	ĩprim'iftə	ĩprim'iu	ĩprim'imuɟ	ĩprim'iftəɟ	ĩprim'irɛũ
FAZER	f'if	fiz'ɛftə	f'ɛɟ	fiz'əmuɟ	fiz'ɛftəɟ	fiz'erɛũ
QUERER	k'if	kiz'ɛftə	k'if	kiz'əmuɟ	kiz'ɛftəɟ	kiz'erɛũ
TRAZER	tr'osə	tros'ɛftə	tr'osə	tros'əmuɟ	tros'ɛftəɟ	tros'erɛũ
OPOR	op'uf	opuz'ɛftə	op'of	opuz'əmuɟ	opuz'ɛftəɟ	opuz'erɛũ
VIR	v'ĩ	vi'ɛftə	v'əiu	vi'əmuɟ	vi'ɛftəɟ	vi'erɛũ

**Tableau 16** Exemples représentatifs de la conjugaison du prétérit perfectif

Un autre exemple du même type se rencontre au présent de l'indicatif. À ce tiroir, les deux seules formes qui sont interprédictibles sont la 2SG et la 3SG. Nous reviendrons plus loin sur la non-interprédictibilité des autres formes. Pour les deux cas qui nous intéressent, on note que la grande majorité des verbes respectent les conditions en (13) : l'accent est sur la voyelle préthématique, et on a un contraste de voyelle thématique à deux niveaux (/ɛ/ vs. /ə/) avant des terminaisons constantes (/ɟ/ ou zéro). Il y a cependant un petit nombre de verbes irréguliers qui utilisent à la 3SG un pseudo-radical nu : d'une part les 14 verbes des 2<sup>e</sup> et 3<sup>e</sup> conjugaisons à radical en /-z/ (le /z/ est palatalisé et dévoisé en fin de mot, sous l'effet la phonologie régulière), d'autre part les deux verbes QUERER et REQUERER, qui se trouvent être les seuls verbes à radical en /-kɛr/. Dans les deux cas, les caractéristiques phonotactiques spécifiques de la sous-classe évitent toute ambiguïté : aucun verbe ne se termine en /ɟ/ ou en /r/ à la 3SG s'il ne relève d'un de ces cas ; de même pour les verbes qui se terminent en /-zəɟ/ ou en /-kɛrəɟ/ à la 2SG.

	1SG	2SG	3SG	1PL	2PL	3PL
utilizar	util'izu	util'izɛɟ	util'izə	utiliz'əmuɟ	utiliz'aif	util'izɛũ
aprender	ɛpr'ɛdu	ɛpr'ɛdɛɟ	ɛpr'ɛdə	ɛprɛd'əmuɟ	ɛprɛd'ɛif	ɛpr'ɛdɛĩ
imprimir	ĩpr'imu	ĩpr'imɛɟ	ĩpr'imə	ĩprim'imuɟ	ĩprim'if	ĩpr'imɛĩ
traduzir	trɛd'uzu	trɛd'uzɛɟ	trɛd'uf	trɛduz'imuɟ	trɛduz'if	trɛd'uzɛĩ
requerer	rək'ɛru	rək'ɛrɛɟ	rək'ɛr	rəkər'əmuɟ	rəkər'ɛif	rək'ɛrɛĩ

**Tableau 17** Le présent de l'indicatif de quelques verbes, illustrant l'interprédictibilité du 2SG et du 3SG.

## IV.2 Distillation du paradigme

À partir de l'observation des relations d'interprédictibilité mutuelle, il est possible de construire des ensembles de cases du paradigme qui sont toutes interprédictibles entre elles. En choisissant les ensembles les plus grands possibles, on aboutit à une *partition* du paradigme en zones de parfaite interprédictibilité. Ces zones sont le pendant direct, dans l'approche à base de mots, de la notion de case de l'espace thématique dans l'approche à base de radicaux de Bonami & Boyé (2003)<sup>16</sup>. Dans le cas présent on aboutit à une partition à 12 cellules, illustrée dans la figure 3<sup>17</sup>. Chaque numéro correspond à une zone d'interprédictibilité.

<sup>16</sup> Voir également les notions de *overall distribution schema* de Pirrelli & Battista (2000) ou d'*alliances of forms* d'Ackerman, Blevins & Malouf (2009).

<sup>17</sup> L'interprédictibilité n'est pas une relation transitive : il peut arriver qu'on trouve 3 cases du paradigme *A*, *B* et *C* telles que *A* et *B* sont interprédictibles, *B* et *C* sont interprédictibles, mais *A* et *C*

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	1	2		3	4	5
PST.IPFV.IND	6					
PST.PFV.IND	7	8				
PST.PRF.IND						
FUT.IND	9					
COND						
PRS.SBJV	10			11		10
PST.SBJV	8					
FUT.SBJV						
IMP	—	2	10	11	4	10
INF.CONJ	3					

INF	GER	PST.PTCP
3		12

**Figure 3** Partition du paradigme verbal du portugais européen induite à partir des relations d'interprédictibilité totale

La partition de la figure 3 fait ressortir le caractère morphomique, au sens d'Aronoff (1994), des zones d'interprédictibilité. Certaines zones correspondent à une classe morphosyntaxiquement naturelle de formes : c'est trivialement le cas des zones singleton (zones 1, 5, 7, 12), c'est clairement le cas pour la zone 6 qui correspond à un tiroir unique, et on pourrait soutenir que c'est également vrai pour la zone 9 (si on soutient que le futur et le conditionnel ont des propriétés sémantiques en commun). En revanche, les inventaires de formes correspondant aux zones restantes ont un caractère clairement disjonctif, comme l'illustre le relevé en (14) : dans chaque cas, il n'y a pas moyen de décrire l'ensemble de cases du paradigme concernées par une description qui ne ferait intervenir que des conjonctions (et pas des disjonctions ou des négations).

- (14) a. Zone 2 : (IND.PRS **et** (2SG **ou** 3SG)) **ou** IMP.2SG  
 b. Zone 3 : INF **ou** GER **ou** PRS.IND.1PL

---

ne le sont pas. De ce fait, il peut exister plus d'une manière de partitionner un paradigme en zones d'interprédictibilité parfaite, même si dans notre expérience cette situation est plutôt exceptionnelle.

Notre méthode de calcul trouve exactement une situation de ce type en portugais : à l'IND.PRS, la 2SG et la 3SG sont interprédictibles, comme le sont le 2SG et la 3PL, mais la 3SG et la 3PL ne le sont pas. Il se trouve que notre programme infère pour la relation [IND.PRS.2SG $\Rightarrow$ IND.PRS.3SG] le patron très spécifique [əf  $\rightarrow$   $\emptyset$  / k \_\_\_ #], alors qu'il infère pour la relation [IND.PRS.3PL $\Rightarrow$ IND.PRS.3SG] le patron plus général [ẽĩ  $\rightarrow$   $\emptyset$  / [+son] \_\_\_ #]. Cela est dû aux verbes tels que TRAIR, qui suivent le cas général entre 2SG et 3SG (suppression du /ʃ/ final) mais sont un cas particulier entre 3PL et 3SG (suppression de la finale /ẽĩ/ sans ajout de /ə/). De ce fait, le verbe QUERER tombe dans la même classe que le verbe PROFERIR pour les besoins de [IND.PRS.3PL $\Rightarrow$ IND.PRS.3SG] mais pas pour celui de [IND.PRS.2SG $\Rightarrow$ IND.PRS.3SG].

Il est difficile de décider si cette situation doit être interprétée comme une propriété remarquable ou comme un artefact dû à la manière d'inférer les patrons d'alternance. Quoi qu'il en soit, il existe de fait une partition alternative à celle du tableau 17, qui groupe PRS.3PL, plutôt que PRS.3SG, avec PRS.2SG.

- c. Zone 4 : (IND.PRS **ou** IMP) **et** 2PL
- d. Zone 8 : (IND.PST **et** (**non** IND.PST.IMP.1SG)) **ou** (SBJV **et** (**non** PRS))
- e. Zone 10 : (SBJV.PRS **et** (SG ou 3PL)) **ou** IMP.3
- f. Zone 11 : (SBJV.PRS **et** (1PL ou 2PL)) **ou** IMP.1PL

Si on se souvient que les zones d'interprédicibilité correspondent presque toujours à des zones de non-allomorphie radicale, on retrouve ici sous une autre forme les observations bien connues sur le caractère morphomique de la distribution des radicaux dans les langues romanes.

À partir de la partition de la figure 3, on peut construire ce que Stump & Finkel (2013) appellent une *distillation du paradigme*. Une distillation est un ensemble de cases dont chaque membre correspond à une zone d'interprédicibilité distincte. L'intérêt d'une distillation est de permettre d'étudier un sous-paradigme de taille nettement plus raisonnable sans perte d'information : étant données deux zones  $z_1$  et  $z_2$ , toute relation implicative d'une case de  $z_1$  à une case de  $z_2$  a un correspondant exact qui relie chaque autre case de  $z_1$  à chaque autre case de  $z_2$ <sup>18</sup>. En principe le choix de la case qui représente chaque zone est sans conséquence ; parmi les 1 609 632 possibilités<sup>19</sup>, nous choisissons celle qui est indiquée dans le Tableau 18.

zone	case	zone	case	zone	case
1	IND.PRS.1SG	5	IND.PRS.3PL	9	FUT.IND.3SG
2	IND.PRS.3SG	6	PST.IPFV.3SG	10	PRS.SBJV.3SG
3	IND.PRS.1PL	7	PST.PFV.1SG	11	PRS.SBJV.2PL
4	IND.PRS.2PL	8	PST.PFV.3SG	12	PST.PTCP

**Tableau 18** Une distillation du paradigme déduite de l'exemple

Une fois que l'on se concentre sur une distillation, il est possible d'examiner en détail les valeurs d'entropie conditionnelle liant les cases du paradigme. Le tableau 19 donne les chiffres détaillés pour la distillation choisie, et le reste de cette section sera consacré à identifier les causes des contrastes de prédictibilité que ce tableau fait ressortir.

<sup>18</sup> L'observation de la note 17 montre que cette assertion peut ne pas être tout à fait exacte, en fonction de la partition utilisée : il peut se trouver que depuis deux cases  $c_1$  et  $c_2$  appartenant à la même cellule de la partition, la prédictibilité d'une case  $c_3$  soit différente, parce que les patrons de  $[c_1 \Rightarrow c_3]$  ne groupent pas de la même manière l'espace de variation phonotactique que ceux de  $[c_2 \Rightarrow c_3]$ . En pratique on peut vérifier automatiquement si cette situation se présente, et elle ne se présente pas pour la partition choisie pour le portugais dans cet article.

<sup>19</sup> Le choix fait parmi l'ensemble des cases de chaque zone est indépendant de ceux faits pour les autres zones. Pour une partition à  $n$  zones  $z_1, \dots, z_n$ , il y a donc  $\prod_{i=1}^n |z_i|$  partitions distinctes.

	PRS.IND.1SG	PRS.IND.3SG	PRS.IND.1PL	PRS.IND.2PL	PRS.IND.3PL	PST.IPFV.3SG	PST.PFV.1SG	PST.PFV.3SG	FUT.IND.3SG	PRS.SBJV.3SG	PRS.SBJV.2PL	PST.PTCP
PRS.IND.1SG	—	0,4864	0,5291	0,5291	0,5246	0,4598	0,4509	0,5291	0,5291	0,8627	0,5114	0,4488
PRS.IND.3SG	0	—	0,0728	0,0728	0	0,0104	0,0104	0,0728	0,0875	0	0,0104	0,0104
PRS.IND.1PL	0,1138	0,1244	—	0,0671	0,1157	0,0227	0,0227	0,0351	0	0,1138	0,0311	0,0157
PRS.IND.2PL	0,1083	0,1063	0,0355	—	0,1063	0,0355	0,0349	0,0349	0,0351	0,1083	0,0065	0,0349
PRS.IND.3PL	0,0085	0,0105	0,0798	0,0798	—	0,0104	0,0104	0,0798	0,0798	0	0,0104	0,0104
PST.IPFV.3SG	0,1161	0,1151	0,2468	0,2601	0,1151	—	0	0,2389	0,2423	0,1161	0,0229	0,0122
PST.PFV.1SG	0,1463	0,1454	0,2886	0,2804	0,1454	0,05	—	0,2856	0,2886	0,1463	0,0699	0,0619
PST.PFV.3SG	0,1063	0,1123	0,0049	0,0194	0,1123	0	0	—	0,0049	0,1063	0	0,0049
FUT.IND.3SG	0,1175	0,142	0,0298	0,0704	0,1327	0,0545	0,0282	0,0406	—	0,1175	0,0378	0,0212
PRS.SBJV.3SG	0	0,0088	0,087	0,087	0,0052	0,008	0,008	0,087	0,087	—	0,0028	0,008
PRS.SBJV.2PL	0,1141	0,1151	0,1962	0,1962	0,1151	0,0088	0,0088	0,1962	0,1962	0,1141	—	0,0088
PST.PTCP	0,125	0,1432	0,2399	0,2512	0,1262	0,0268	0,0268	0,254	0,239	0,125	0,0321	—

**Tableau 19** Entropie conditionnelle pour chaque couple de cases de la distillation du tableau 18

Dans ce tableau, la teinte du fond de chaque case fait ressortir la hauteur relative de la valeur d'entropie trouvée par le programme : plus le fond est foncé, plus la valeur est élevée. Il ressort très clairement au simple examen de ces chiffres que l'IND.PRS.1SG est le plus mauvais prédicteur du reste du paradigme : les valeurs de la première ligne sont plus hautes que toutes les autres valeurs du tableau. Les meilleurs prédicteurs sont au nombre de 3 avec des valeurs moyennes très proches : l'IND.PRS.3SG (moyenne 0,0316), l'IND.PRS.3PL (moyenne 0,0346) et le SBJV.PRS.3SG (moyenne 0,0353). Si l'on examine maintenant le tableau en colonnes, la structure ne ressort pas de manière aussi nette. Il est clair cependant que la difficulté de prédiction se répartit de manière non uniforme : suivant le prédicteur, ce n'est pas la même partie du paradigme qui est difficile à prédire. Dans les paragraphes qui suivent, nous identifions trois causes principales des difficultés de prédiction et montrons comment l'aspect du tableau 19 est expliqué par ces causes. Avant de procéder à cet examen, il est important de rappeler que les chiffres du tableau 19 doivent être interprétés en connaissance des limites de l'ensemble de données examiné et de la procédure d'analyse automatique. Le fait de se focaliser sur les 1000 lexèmes les plus fréquents permet d'avoir un inventaire à peu près exhaustif des patrons de conjugaison (s'il peut exister des patrons concernant seulement des verbes rares, il est très improbable qu'ils concernent un grand nombre de verbes), mais en revanche la distribution statistique des patrons serait différente sur un ensemble plus grand. De plus, la procédure d'identification des patrons d'alternance donne parfois des résultats contre-intuitifs et peu informatifs<sup>20</sup>, qui ne peuvent être identifiés qu'en examinant en détail

<sup>20</sup> Un cas particulièrement net de non-informativité est le suivant : la valeur d'entropie la plus haute dans le tableau concerne le cas de la relation [IND.PRS.1SG⇒SBJV.PRS.3SG]. Ce résultat est surprenant : a priori, le SBJV.PRS.3SG paraît plus facile à prédire à partir de l'IND.PRS.1SG que, par exemple, l'IND.PRS.1PL, parce que le SBJV.PRS.3SG ne contraste que deux voyelles thématiques alors que l'IND.PRS.1PL en contraste trois (voir ci-dessous VI.3).

Ce résultat contre-intuitif est dû aux limites de l'algorithme d'identification des patrons, déjà noté dans la section III.3. Parce que l'accent se déplace vers la voyelle thématique, dans la relation [IND.PRS.1SG⇒IND.PRS.1PL], l'algorithme identifie autant de patrons qu'il y a de triplets voyelle



l'inventaire des patrons identifiés. Dans ce qui suit nous nous contentons de commenter des cas absolument clairs, où les prédictibilités et imprédictibilités notées par l'automate semblent plausibles au linguiste.

### IV.3 Première cause : la neutralisation des conjugaisons

La première cause d'imprédictibilité dans les paradigmes est clairement la neutralisation dans certaines cases des distinctions entre les conjugaisons. Revenons sur l'indicatif présent des verbes réguliers (tableau 4). Les formes de 1PL et 2PL sont les seules à présenter une marque distinctive permettant d'identifier la conjugaison, sous la forme d'une opposition entre 3 voyelles thématiques : /ɐ/ vs. /e/ vs. /i/ à la 1PL, /ai/ vs. /ɛi/ vs. /i/ à la 2PL. Trois autres formes neutralisent la distinction entre 2<sup>e</sup> et 3<sup>e</sup> conjugaison : on a /ɐ/ vs. /ə/ aux 2SG et 3SG, /ẽũ/ vs. /ẽĩ/ à la 3PL. Il suit de cette observation que la prédiction de la 1PL à partir du 3SG par exemple pose un problème : si la 2SG se termine par /ə/, on ne sait pas s'il faut rétablir /e/ ou /i/ à la 1PL. Dans la direction inverse, la prédiction de la voyelle thématique ne pose pas de difficulté. La forme de 1SG est plus problématique encore puisqu'elle neutralise les distinctions entre les trois conjugaisons. Selon le même raisonnement, on s'attend donc à avoir des difficultés à prédire les autres formes : pour prédire la 3SG, on ne sait pas s'il faut rétablir /ɐ/ ou /ə/ ; pour la 1PL, s'il faut rétablir /ɐ/, /e/ ou /i/.

Toutes choses égales par ailleurs, ces observations amènent à présumer qu'à l'indicatif présent, la prédictibilité devrait être meilleure du groupe de formes {2SG, 3SG, 3PL} vers {1PL, 2PL} que dans le sens inverse ; de même, on s'attend à ce que la prédiction à partir de la 1SG soit plus difficile que la prédiction de la 1SG, et à ce qu'à partir de la 1SG, il soit plus facile de prédire les formes du groupe {2SG, 3SG, 3PL} que celles du groupe {1PL, 2PL}. Le lecteur peut vérifier dans le tableau 19 que toutes ces prédictions sont vérifiées. Particulièrement frappant est le fait que la 1SG est catégoriquement prédictible à partir de la 3SG, mais que l'inverse est loin d'être vrai.

Les observations qui viennent d'être faites concernant l'indicatif présent se généralisent à l'ensemble du paradigme : prédire une case qui fait plus de distinctions entre classes flexionnelles à partir d'une case qui en fait moins est toujours plus difficile que l'inverse. L'influence majeure de ce facteur sur les valeurs d'entropie du tableau 19 est mise en avant quand on compare celui-ci au tableau 20, qui explicite pour chaque couple de cases de la distillation, le nombre de distinctions que chacune des deux cases met en œuvre. On constate que sur les 131 couples de cases de la distillation, 39 mettent en jeu un couple dont le second élément fait plus de distinctions de voyelles thématiques que le premier. Sur ces 39 cases, 27 ont une valeur d'entropie supérieure à 0,15 ; aucune autre case du paradigme n'a de valeur supérieure. De plus, dans chaque ligne et dans chaque colonne, les couples qui correspondent à une distinctivité croissante ont une valeur d'entropie supérieure aux autres. Par exemple à partir de l'imparfait 3SG, les formes les plus difficiles à prédire sont celles qui distinguent

---

préthématique-séquence de consonnes-voyelle thématique. Parce qu'il est inféré à partir d'un petit nombre de verbes, chacun de ces patrons est associé à un contexte phonotactique étroit, et introduit donc peu, voire pas, d'incertitude ; d'où une valeur d'entropie relativement basse. À l'inverse, dans la relation [IND.PRS.1SG⇒SBJV.PRS.3SG], l'accent est fixe, et deux patrons globaux, correspondant aux deux choix de voyelle thématique, sont inférés. Chacun des deux patrons est associé à un contexte phonotactique maximalement général, et l'incertitude est donc maximale.

On voit ici comment des choix analytiques qui sont motivés dans le cas général peuvent donner des résultats peu utiles dans certains cas particuliers. La solution à ce problème n'est sans doute pas de remplacer la mesure de prédictivité utilisée ici, mais de multiplier les mesures.

trois voyelles thématiques ; à l'inverse, les formes qui distinguent 3 voyelles thématiques sont toutes de meilleurs prédicteurs de l'IND.PRS.1PL que celles qui n'en distinguent que 2.

	PRS.IND.1SG	PRS.IND.3SG	PRS.IND.1PL	PRS.IND.2PL	PRS.IND.3PL	PST.IPFV.3SG	PST.PFV.1SG	PST.PFV.3SG	FUT.IND.3SG	PRS.SBJV.3SG	PRS.SBJV.2PL	PST.PTCP
PRS.IND.1SG	—	1→2	1→3	1→3	1→2	1→2	1→2	1→3	1→3	1→2	1→2	1→2
PRS.IND.3SG	2→1	—	2→3	2→3	2→2	2→2	2→2	2→3	2→3	2→2	2→2	2→2
PRS.IND.1PL	3→1	3→2	—	3→3	3→2	3→2	3→2	3→3	3→3	3→2	3→2	3→2
PRS.IND.2PL	3→1	3→2	3→3	—	3→2	3→2	3→2	3→3	3→3	3→2	3→2	3→2
PRS.IND.3PL	2→1	2→2	2→3	2→3	—	2→2	2→2	2→3	2→3	2→2	2→2	2→2
PST.IPFV.3SG	2→1	2→2	2→3	2→3	2→2	—	2→2	2→3	2→3	2→2	2→2	2→2
PST.PFV.1SG	2→1	2→2	2→3	2→3	2→2	2→2	—	2→3	2→3	2→2	2→2	2→2
PST.PFV.3SG	3→1	3→2	3→3	3→3	3→2	3→2	3→2	—	3→3	3→2	3→2	3→2
FUT.IND.3SG	3→1	3→2	3→3	3→3	3→2	3→2	3→2	3→3	—	3→2	3→2	3→2
PRS.SBJV.3SG	2→1	2→2	2→3	2→3	2→2	2→2	2→2	2→3	2→3	—	2→2	2→2
PRS.SBJV.2PL	2→1	2→2	2→3	2→3	2→2	2→2	2→2	2→3	2→3	2→2	—	2→2
PST.PTCP	2→1	2→2	2→3	2→3	2→2	2→2	2→2	2→3	2→3	2→2	2→2	—

**Tableau 20** Influence de la distinctivité des voyelles thématiques sur la prédictibilité<sup>21</sup>

#### IV.4 Deuxième cause : les réductions vocaliques

Si les contrastes de distinctivité des voyelles thématiques sont le facteur qui a la plus grande influence sur la prédictibilité en portugais, ce n'est bien évidemment pas le seul. Un autre facteur majeur est la réduction des voyelles inaccentuées. En portugais européen, les voyelles orales inaccentuées se réduisent selon un patron qui induit des neutralisations de distinctions visibles seulement sous accent (voir par exemple Mateus & d'Andrade (2000 : 17-23)<sup>22</sup>. Le tableau 21 inventorie les alternances en question.

accentué	i	e	ɛ	a	ɐ	ɔ	o	u
inaccentué	i	ə		ɐ			u	

**Tableau 21** Réduction des voyelles orales inaccentuées en portugais européen

Bien qu'ils relèvent entièrement de la phonologie régulière, ces contrastes ont des conséquences importantes sur les relations de prédictibilité entre formes. Une fois de plus, concentrons-nous sur l'indicatif présent pour commencer. Aux 1SG, 2SG, 3SG et 3PL, l'accent tombe sur la voyelle préthématique. Aux 1PL et 2PL, il tombe sur la voyelle thématique. De ce fait, la voyelle préthématique se trouve inaccentuée, ce qui conduit à des neutralisations, qui sont illustrées dans le tableau 22 avec des verbes de la 1<sup>e</sup> conjugaison.

<sup>21</sup>  $m \rightarrow n$  signifie que la case de départ distingue  $m$  marqueurs de classe et que la classe d'arrivée en distingue  $n$ .

<sup>22</sup> En fonction de la variété et du débit, certaines voyelles inaccentuées peuvent ne pas être réalisées. On ignorera ce phénomène pour les besoins de cet article.

	1SG	2SG	3SG	1PL	2PL	3PL
UTILIZAR	util'izu	util'izɐʃ	util'ize	utiliz'ɐmuʃ	utiliz'aiʃ	util'izẽũ
ENCENAR	ẽs'enu	ẽs'enɐʃ	ẽs'enɛ	ẽsɛn'ɐmuʃ	ẽsɛn'aiʃ	ẽs'enẽũ
ENCETAR	ẽs'ɛtu	ẽs'ɛtɐʃ	ẽs'ɛtɛ	ẽsɛt'ɐmuʃ	ẽsɛt'aiʃ	ẽs'ɛtẽũ
MATAR	m'atu	m'atɐʃ	m'atɛ	mɐt'ɐmuʃ	mɐt'aiʃ	m'atẽũ
CHAMAR	ʃ'ɛmu	ʃ'ɛmɐʃ	ʃ'ɛmɛ	ʃɛm'ɐmuʃ	ʃɛm'aiʃ	ʃ'ɛmẽũ
CONVOCAR	kõv'ɔku	kõv'ɔkɐʃ	kõv'ɔkɛ	kõvuk'ɐmuʃ	kõvuk'aiʃ	kõv'ɔkẽũ
TELEFONAR	tələf'onu	tələf'onɐʃ	tələf'onɛ	tələfun'ɐmuʃ	tələfun'aiʃ	tələf'onẽũ
EDUCAR	ed'uku	ed'ukɐʃ	ed'ukɛ	eduk'ɐmuʃ	eduk'aiʃ	ed'ukẽũ

**Tableau 22** Effet des réductions vocaliques sur l'indicatif présent

Parce qu'elles créent de l'opacité, les réductions vocaliques créent de l'imprédictibilité : face à une 1PL ou 2PL à voyelle préthématique /ə/, il existe une incertitude quant à la qualité de la préthématique des autres formes, qui peut être /e/ ou /ɛ/ ; de même pour les 1PL et 2PL à préthématiques en /ɐ/ et en /u/. Le potentiel d'incertitude est loin d'être négligeable : seuls les verbes qui ont une voyelle préthématique /i/ ou nasale en sont indemnes. Cependant l'incertitude avérée n'est pas très élevée étant donnée la distribution des voyelles accentuées. Le tableau 23 indique, pour les 494 verbes réguliers<sup>23</sup> présentant une voyelle neutralisée à la 1PL, la distribution des voyelles à la 1SG.

voyelle préthématique du 3SG	e	ɛ	a	ɐ	ɔ	o	u
voyelle préthématique du 1PL	ə	ə	ɐ	ɐ	u	u	u
effectif de la classe	10	168	95	12	116	6	87

**Tableau 23** Classification des verbes réguliers en fonction des alternances de voyelle préthématique à l'indicatif présent

On voit que pour les verbes à 1PL en /ə/ et en /ɐ/, la distribution est très biaisée ; de ce fait, si la prédiction de la voyelle pleine à partir de la voyelle réduite ne peut être faite de manière catégorique, le niveau d'incertitude est faible. En revanche, pour les verbes à 1PL en /u/, la distribution est équilibrée entre les voyelles pleines /ɔ/ et /u/, et le niveau d'incertitude est donc assez élevé pour ces verbes, qui constituent environ un quart du lexique. Le fait que seul un quart du lexique soit concerné par une incertitude notable explique que les valeurs d'entropie de l'ensemble de cases {1PL, 2PL} vers l'ensemble de cases {1SG, 2SG, 3SG, 3PL}, quoi que non-négligeables, soient relativement basses (de l'ordre de 0,1) ; en particulier, l'incertitude résultant des réductions vocaliques est nettement plus faible que celle qui résulte de la neutralisation des voyelles thématiques.

Les observations qui viennent d'être faites peuvent être étendues à l'ensemble de la distillation. Le tableau 24 indique pour chaque couple de formes l'alternance de placement de l'accent. La notation « = » indique que l'accent est dans la même position dans les deux cases ; « ↓ » qu'il se déplace depuis la voyelle préthématique vers une autre voyelle ; « ↑ » qu'il se déplace vers la voyelle préthématique. C'est dans ce dernier cas que les réductions vocaliques produisent de l'incertitude. En examinant le tableau, on constate que les cases correspondantes (entourées en lignes pleines) sont celles qui ont les valeurs d'entropie les plus élevées, une fois qu'on met de côté les cas expliqués par les neutralisations de voyelles thématiques (entourés en lignes pointillées). On a donc clairement identifié la seconde source d'imprédictibilité dans la conjugaison du portugais européen.

<sup>23</sup> On a laissé de côté les verbes à voyelle préthématique /i/ ou nasale, qui n'alternent pas, et les verbes ne présentant pas de terminaison régulière au 3SG.

	PRS.IND.1SG	PRS.IND.3SG	PRS.IND.1PL	PRS.IND.2PL	PRS.IND.3PL	PST.IPFV.3SG	PST.PFV.1SG	PST.PFV.3SG	FUT.IND.3SG	PRS.SBJV.3SG	PRS.SBJV.2PL	PST.PTCP
PRS.IND.1SG	—	=	↓	↓	=	↓	↓	↓	↓	=	↓	↓
PRS.IND.3SG	=	—	↓	↓	=	↓	↓	↓	↓	=	↓	↓
PRS.IND.1PL	↑	↑	—	=	↑	=	=	=	=	↑	=	=
PRS.IND.2PL	↑	↑	=	—	↑	=	=	=	=	↑	=	=
PRS.IND.3PL	=	=	↓	↓	—	↓	↓	↓	↓	=	↓	↓
PST.IPFV.3SG	↑	↑	=	=	↑	—	=	=	=	↑	=	=
PST.PFV.1SG	↑	↑	=	=	↑	=	—	=	=	↑	=	=
PST.PFV.3SG	↑	↑	=	=	↑	=	=	—	=	↑	=	=
FUT.IND.3SG	↑	↑	=	=	↑	=	=	=	—	↑	=	=
PRS.SBJV.3SG	=	=	↓	↓	=	↓	↓	↓	↓	=	↓	↓
PRS.SBJV.2PL	↑	↑	=	=	↑	=	=	=	=	↑	—	=
PST.PTCP	↑	↑	=	=	↑	=	=	=	=	↑	=	—

**Tableau 24** Influence de la place de l'accent sur la prédictibilité

L'interaction entre les deux causes d'imprédictibilité identifiées jusqu'ici donne lieu à une situation remarquable. Il existe une corrélation entre la place de l'accent et le nombre de distinctions de voyelles thématiques : les formes qui distinguent 3 voyelles thématiques sont celles qui sont accentuées sur celle-ci (ou sur la post-thématique dans le cas du futur et du conditionnel) ; ce sont donc aussi celles qui ont une voyelle préthématique inaccentuée. De ce fait, l'ensemble des couples de cases où une incertitude est introduite par les voyelles thématiques est disjoint de l'ensemble de couples où elle est introduite par les alternances de voyelles préthématiques.

Cette situation a des répercussions sur la possibilité de définir un système de parties principales pour la conjugaison du portugais européen. À la suite de Finkel & Stump (2007), on définit un système de parties principales comme un ensemble de cases du paradigme à partir duquel toutes les autres cases peuvent être déduites de manière catégorique. Dans le cas idéal, un système flexionnel possède une partie principale unique (voir la *Uniform Base Hypothesis* d'Albright (2002)). Cette unique partie principale est un bon candidat à servir de forme de citation, dans la mesure où sa connaissance suffit à connaître l'ensemble du paradigme. En pratique cependant, comme le montrent en détail Stump & Finkel (2013), il est rare qu'une partie principale suffise, ne serait-ce que parce que certains lexèmes irréguliers ont une flexion imprédictible.

L'interaction entre neutralisations de conjugaison et réductions vocaliques a pour conséquence que même si on se limite au sous-système constitué par les verbes réguliers des trois conjugaisons traditionnelles, au moins deux parties principales sont nécessaires : toute forme à accent sur la voyelle préthématique sera un mauvais prédicteur des formes à accent sur la voyelle thématique ou post-thématique, et réciproquement.

#### IV.5 Troisième cause: irrégularités locales

Les deux causes déjà identifiées rendent compte des 56 cases du tableau 19 ayant les valeurs d'entropie les plus élevées. Dans les cases restantes, l'entropie est partout inférieure à 0,075. La cause majeure de l'incertitude restante est l'existence de phénomènes d'allomorphie non réguliers. Une fois encore, on peut illustrer la situation générale en prenant des exemples issus de l'indicatif présent. Prenons la relation entre [IND.PRS.1PL⇒IND.PRS.2PL]. L'ensemble des patrons d'alternance rencontré est indiqué dans le tableau 25.

patron	effectif	alternance radicale
ɐmuʃ → aɪʃ / __#	704	X~X
ɛmuʃ → ɛɪʃ / __#	130	X~X
imuʃ → iʃ / __#	138	X~X
omuʃ → ɔ̃dɔʃ / p__#	13	Xv~X̃vdə
ɛmuʃ → edɔʃ / v__#	4	Xv~X̃vdə
imuʃ → idɔʃ / r__#	2	Xv~X̃vdə
ɛmuʃ → ẽdɔʃ / t__#	8	Xv~X̃vdə
imuʃ → ïdɔʃ / v__#	2	Xv~X̃vdə
muʃ → iʃ / #so__#	1	so~soi

**Tableau 25** Inventaire des patrons d'alternance mis en œuvre dans la relation [IND.PRS.1PL⇒IND.PRS.2PL]

On voit qu'il existe un petit ensemble de verbes qui ne suivent aucun des trois patrons réguliers des trois conjugaisons. Étant donné les hypothèses de segmentation esquissées dans la section III.1, il est naturel d'interpréter cette situation comme un cas d'allomorphie radicale : à la stratégie qui utilise le même radical à la 1PL et à la 2PL s'opposent deux autres stratégies, qui utilisent à la 2PL un radical augmenté de /də/, avec ou sans nasalisation de la voyelle thématique ; le verbe SER présente une alternance allomorphique isolée.

Cette situation d'allomorphie radicale conduit à une certaine incertitude, dans la mesure où certains patrons d'alternance sont applicables dans les mêmes contextes phonologiques. Le tableau 26 fait l'inventaire de ces situations. Cependant, l'effectif des classes concernées est trop faible pour faire une contribution forte à l'entropie : au total 49 verbes sont concernés, et la distribution des deux patrons concernés est loin d'être équilibrée dans la plupart des cas.

classe	exemple	effectif	patron	effectif
c <sub>1</sub>	HAYER,	22	ɛmuʃ → ɛɪʃ / __#	18
	VER		ɛmuʃ → edɔʃ / [+cons,+voix]__#	4
c <sub>2</sub>	RIR	2	imuʃ → iʃ / __#	0
			imuʃ → idɔʃ / r__#	2
c <sub>3</sub>	BEBER,	21	ɛmuʃ → ɛɪʃ / __#	13
	TER		ɛmuʃ → ẽdɔʃ / t__#	8
c <sub>4</sub>	OUVIR,	4	imuʃ → iʃ / __#	2
	VIR		imuʃ → ïdɔʃ / v__#	2

**Tableau 26** Classes d'IND.PRS.1PL à deux patrons dans la relation [IND.PRS.1PL⇒IND.PRS.2PL]

De manière remarquable, l'allomorphie radicale ne cause pas d'incertitude dans la direction [IND.PRS.2PL⇒IND.PRS.1PL], la forme du radical « court » étant entièrement prédictible sur la base de celle du radical long<sup>24</sup>.

La relation [IND.PRS.3PL⇒IND.PRS.3SG] illustre un dernier type de situation d'incertitude, due cette fois-ci à des terminaisons irrégulières et non à de l'allomorphie radicale. Les verbes des 2<sup>e</sup> et 3<sup>e</sup> conjugaison se terminent normalement en /ẽĩ/ à la 3PL et en /ə/ à la 3SG. Il existe cependant quelques verbes qui n'ont pas la terminaison attendue à la 3SG, ce que nous avons déjà noté en commentant le tableau 5. Dans la plupart des cas il n'en résulte pas d'incertitude, parce que la classe de verbes concernée a un aspect phonologique distinctif : ainsi tous les verbes à 3PL en /zẽĩ/ ont une 3SG en /ʃ/. Le seul cas qui crée de l'incertitude est celui des verbes QUERER et REQUERER : ils se terminent à la 3PL en /rẽĩ/, et à la 3SG en /r/, alors que 27 autres verbes ayant une 3PL de même aspect ont une 3SG régulière en /rə/. La quantité d'incertitude est cependant très faible (entropie conditionnelle 0,0105), puisque d'une part le nombre de verbes concernés est très faible (29), et d'autre part la distribution des deux patrons est très inégale (plus de 90% des cas relèvent du patron régulier)<sup>25</sup>.

Il n'est pas question ici de commenter une par une les causes de chacune des valeurs d'entropie trouvées pour les 74 couples de cases de la distillation qui n'ont pas encore été mentionnés ; on se contentera de noter que les valeurs trouvées sont dues à diverses combinaisons des facteurs déjà identifiés.

## V. Comparaison entre le portugais européen et le français

Sans avoir rendu compte de la totalité des causes de l'incertitude dans la conjugaison du portugais, les sections qui précèdent ont donné un certain nombre d'exemples qui illustre la typologie des causes rencontrées. Celles-ci sont rappelées en (15).

- (15)a. **Neutralisation des conjugaisons** : la case *c* utilise un inventaire de marqueurs de conjugaison (par exemple de voyelles thématiques) plus petit que la classe *c'*.
- b. **Neutralisation phonologique** : un phénomène phonologique régulier aboutit à la neutralisation dans la case *c* d'une opposition phonologique visible dans la case *c'*.
- c. **Allomorphie radicale imprédictible** : l'alternance entre la case *c* et la case *c'* met en jeu une allomorphie radicale pour certains lexèmes, et la présence de l'allomorphie n'est pas prédictible à partir de l'aspect du radical utilisé dans la case *c*.
- d. **Terminaison irrégulière** : certains lexèmes ont une terminaison exceptionnelle pour la case *c'*, et l'occurrence de cette terminaison exceptionnelle n'est pas prédictible à partir de la forme remplissant la case *c*.

---

<sup>24</sup> La valeur d'entropie non nulle notée dans le tableau 19 pour [IND.PRS.2PL⇒IND.PRS.1PL] est due à un défaut de méthode identifié tardivement. Le lexique utilisé code par la même symbole pour la semi-voyelle /j/ et la voyelle /i/ dans la mesure où la distinction est prédictible à partir de l'accent : /i/ non accentué après une autre voyelle est semi-vocalisé. Cependant, l'information sur l'accent est ignorée par l'algorithme d'inférence des patrons d'alternance. De ce fait, les verbes de la 3<sup>e</sup> conjugaison à radical en /-v/ (comme CONTRAIR) ne se distinguent pas à la 2PL des verbes de la 2<sup>e</sup> conjugaison à radical consonantique (comme SOFRER) : dans les deux cas la 2PL se termine en /Cœj/. Si on rétablissait dans les transcriptions l'opposition entre /j/ et /i/, l'entropie serait nulle.

<sup>25</sup> L'existence de terminaisons irrégulières à la 3SG ne cause pas d'incertitude dans la direction [IND.PRS.3SG⇒IND.PRS.3PL], la forme du 3PL n'étant en rien exceptionnelle ; à partir d'une forme irrégulière, prédire la forme régulière ne pose pas de difficulté. L'entropie est donc de 0 dans cette direction.

Si on compare la situation relevée en portugais avec la situation du français telle qu'elle est décrite par Bonami & Boyé (sous presse), il est remarquable de constater que les mêmes facteurs se rencontrent, mais dans des proportions différentes. En français, la source principale d'imprédictibilité est l'interaction de deux phénomènes de neutralisation phonologique : la dégémination de /j/ et la diérèse après attaque branchante. Ces deux phénomènes concourent à faire des formes de l'imparfait et du subjonctif 1PL et 2PL de mauvais prédicteurs du reste du paradigme. Le tableau 27 illustre ce point : un imparfait 2PL se terminant en /CLije/ peut correspondre aussi bien à un radical en /CLij/ avec dégémination de la séquence de deux /j/ produite par l'adjacence avec le marqueur /j/, qu'à un radical en /CL/ avec insertion de /i/ avant le marqueur /j/. Il en résulte que même des formes basées sur le même radical, comme celles de l'imparfait 3SG, sont difficilement prédictibles.

lexème	imparfait 1PL	segmentation	imparfait 3SG
CADRER	kadʁije	kadʁ+j+e	kadʁe
QUADRILLER	kadʁije	kadʁij+j+e	kadʁije

**Tableau 27 Un exemple de neutralisation phonologique dans la conjugaison du français**

La neutralisation des conjugaisons ne joue qu'un rôle limité en français, dans la mesure où peu de cases du paradigme manifestent un marquage différentiel de la classe de conjugaison : essentiellement l'infinitif, le passé simple, l'imparfait du subjonctif, et le participe passé ; et là où ce marquage existe, la prépondérance de la première conjugaison diminue son incidence. En revanche, l'allomorphie radicale joue un rôle majeur, à cause principalement des alternances  $X_i \sim X_{is}$  dans la deuxième conjugaison. De ce fait, une forme en /i/ au présent singulier peut correspondre à une forme en /je/ ou /ise/ à la 2PL (PALLIER vs. PALIR : *je pallie/pâlis, nous pallions, nous pâlissons*) ; à l'inverse, une forme en /ise/ à la 2PL peut correspondre à une forme en /i/ ou en /is/ au singulier (TAPIR vs. TAPISSER : *vous tapissez, je tapis, je tapissez*). Des patrons d'allomorphie radicale plus locaux, comme les alternances  $X \sim XC$ , font néanmoins une contribution importante à l'imprédictibilité : pour aller d'une forme sans consonne à une forme avec consonne, l'incertitude est augmentée par le nombre élevé de possibilités dans le choix de la consonne<sup>26</sup>. Enfin, il existe de très rares cas de terminaisons irrégulières imprédictibles, principalement les 2PL en /t/ de DIRE et FAIRE et de certains membres de leurs familles morphologiques respectives.

De ces observations, il ressort clairement que les types de facteurs identifiés en (15) jouent également un rôle dans la conjugaison du français, mais qu'ils le font dans des proportions très différentes, et affectent des zones différentes du système : le français est dominé par l'allomorphie radicale, alors que le portugais est dominé par le marquage des oppositions de conjugaison par voyelle thématique.

<sup>26</sup> Le français illustre un cinquième type de cause d'imprédictibilité, dont nous n'avons pas trouvé d'exemple clair en portugais. Quand un verbe se termine à l'infinitif en *-ir*, le *i* peut être le premier segment de la marque de l'infinitif (sortir) ou le dernier segment du radical (finir). De ce fait, si on essaye de prédire par exemple le présent 3SG à partir de l'infinitif, il y a une incertitude quant au maintien ou non de la voyelle en question. On peut parler dans ce cas d'*ambiguïté de segmentation* : une des terminaisons *t* pour la case *c* est une sous-chaîne d'une autre terminaison possible *t'*, si bien que certaines formes remplissant la case *c* sont susceptibles de deux segmentations. L'incertitude résulte du fait que chacune de ces segmentations correspond à une possibilité distincte pour la case *c'*.

## VI. Conclusions

Cet article a été consacré à la présentation et à l'illustration d'une méthode d'analyse de la structure implicative des systèmes flexionnels. La méthode repose sur cinq étapes détaillées en (16) :

- (16) a. Constitution d'un large lexique de formes fléchies transcrites phonétiquement.
- b. Construction d'hypothèses formalisées sur les types de patrons d'alternances entre formes exhibés par le lexique en question.
- c. Implémentation d'un programme inférant automatiquement des patrons d'alternance à partir du lexique.
- d. Etude automatique de la distribution statistique des patrons d'alternance.
- e. Interprétation critique des résultats obtenus.

L'étude menée ici sur le portugais européen, si elle a un caractère préliminaire, illustre les deux vertus évidentes de cette méthode de travail : d'une part, elle permet de construire des généralisations qui mettent en jeu de trop gros ensembles de données pour pouvoir être obtenues en examinant le lexique à l'œil nu ; d'autre part, elle permet de quantifier l'importance relative des phénomènes identifiés. Il est cependant important de garder à l'esprit le fait que le caractère quantitatif et partiellement automatisé du travail ne signifie pas qu'il ne doit pas être évalué selon les critères habituels pour une description morphologique. De fait, les choix de transcription faits à l'étape (16a) et les hypothèses sur les patrons d'alternances formulées à l'étape (16b) sont des hypothèses linguistiques d'un caractère familier, et elles conditionnent les résultats de l'analyse tout aussi concrètement ici que dans un travail de description traditionnel.

Le travail mené ici est essentiellement un travail descriptif : il consiste à mettre au jour et à quantifier des régularités dans la structure implicative des paradigmes de conjugaison, et constitue ainsi un ensemble de données dont les théories de la flexion devraient s'attacher à rendre compte. En tant que tel il n'est donc pas lié à une hypothèse théorique particulière. Cependant, par l'importance et la diversité des effets implicatifs qu'il met à jour, ce travail est congruent avec les approches théoriques de la flexion qui mettent l'accent sur les relations entre mots plutôt que sur leur structure morphotactique (par exemple Ford & Singh, 1997 ; Blevins, 2006, sous presse ; Ackerman, Blevins & Malouf, 2009, Ackerman & Malouf, 2013). De ce point de vue, il illustre la nécessité, pour aller de l'avant, de progresser dans la compréhension de la typologie des patrons d'alternance et d'établir un langage de description suffisamment riche pour ces derniers. La pleine prise en compte de la morphologie implicative passe par ce préalable.

## Bibliographie

- ACKERMAN, F., BLEVINS, J. P., & MALOUF, R., 2009. « Parts and wholes: implicative patterns in inflectional paradigms », in J. P. Blevins, J. Blevins (eds.), *Analogy in Grammar*. Oxford, Oxford University Press, 54–82.
- ACKERMAN, F., & MALOUF, R., 2013. « Morphological organization: the low conditional entropy conjecture ». *Language*, 89, 429-464.
- ALBRIGHT, A. C., 2002. *The Identification of Bases in Morphological Paradigms*. Thèse de l'Université de Californie-Los Angeles.
- ARONOFF, M., 1994. *Morphology by itself*. Cambridge, MIT Press.



- BAERMAN, M., BROWN, D., & CORBETT, G. G., 2005. *The Syntax-Morphology Interface: A Study of Syncretism*. Cambridge, Cambridge University Press.
- BLEVINS, J. P., 2006. « Word-based morphology ». *Journal of Linguistics*, 42, 531-573.
- BLEVINS, J., à paraître. *Word and Paradigm Morphology*. Oxford, Oxford University Press.
- BONAMI, O. & BOYE, G., 2003. « Supplétion et classes flexionnelles dans la conjugaison du français », *Langages*, 152:102–126.
- BONAMI, O. & BOYE, G., 2007. « Remarques sur les bases de la conjugaison », in E. Delais-Roussarie et L. Labrune (eds.), *Des sons et des sens*. Paris, Hermès, 77–90.
- BONAMI, O. & BOYE, G., sous presse. « De formes en thèmes », in F. Villoing, S. Leroy, S. David (eds.), *Du régulier et du minutieux*, Nanterre, Presses de l'Université de Paris Ouest.
- BONAMI, O., BOYE, G., GIRAUDO, H., & VOGA, M., 2008. « Quels verbes sont réguliers en français ? », *Actes du premier Congrès Mondial de Linguistique Française*, 1511–1523.
- BOYE, G., 2006. « Suppletion », in K. Brown (ed.), *Encyclopedia of Language and Linguistics*, Elsevier, 297-299.
- BOYE, G. & CABREDO HOFHERR, P., 2006. « The structure of allomorphy in spanish verbal inflection », *Cuadernos de Lingüística*, 13, 9-24.
- BOYE, G., 2000. *Problèmes de morpho-phonologie verbale en français, en espagnol et en italien*. Thèse de doctorat, Université Paris 7.
- BOYE, G., 2011. « Régularités et classes flexionnelles dans la conjugaison du français », in M. Roché, G. Boyé, N. Hathout, S. Lignon & M. Plénat (eds.), *Des unités morphologiques au lexique*, Paris, Hermès, 41-68.
- CORBETT, G., 2007. « Canonical typology, suppletion and possible words ». *Language*, 83, 8-42.
- DELL, F., 1985. *Des règles et des sons*, deuxième édition. Paris, Hermès.
- ESHER, L., 2012. *Future, Conditional, and Autonomous Morphology in Occitan*. Thèse de l'U. d'Oxford.
- FINKEL, R. & STUMP, G., 2007. « Principal parts and morphological typology », *Morphology*, 17, 39-75.
- FORD, A., SINGH, R., & MARTOHARDJONO, G., 1997. *Pace Panini: Towards a Word-Based Theory of Morphology*, New York, Peter Lang.
- GUERRERO, A., 2011. « Verbal inflection in Central Catalan: a realisation analysis », *Lingue e linguaggio*, 2011, 265–282.
- LAVALLEE, J.-F. & LANGLAIS, P., 2011. « Moranapho: un système multilingue d'analyse morphologique basé sur l'analogie formelle », *TAL*, 52(2), 17-44.
- MAIDEN, M., 1992. « Irregularity as a determinant of morphological change ». *Journal of Linguistics* 28, 285–312.
- MAIDEN, M., 2005. « Morphology autonomy and diachrony », in G. Booij & J. Van Marle (eds.), *Yearbook of Morphology 2004*, Dordrecht, Kluwer, 173-175.
- MAIDEN M., 2009. « From pure phonology to pure morphology. The reshaping of the Romance verb », *Revue de linguistique de Vincennes* 38, 45-82.
- MALOUF, R. & ACKERMAN, F., à paraître. « Information-Theoretic Morphology: The Low Entropy Conjecture », *Language*.
- MATEUS, M. H. & D'ANDRADE, E., 2000. *The phonology of Portuguese*, Oxford, Oxford University Press.
- MATTHEWS, P., 1972. *Inflectional morphology: A theoretical study based on aspects of Latin verb conjugation*, Cambridge, Cambridge University Press.
- MONTERMINI, F. & BONAMI, O., 2013. « Stem spaces and predictability in verbal inflection », *Lingue e linguaggio*, XII, 171-190.

- MONTERMINI, F. & BOYE, G., 2012. « Stem relations and inflection class assignment in Italian », *Word Structure* 5, 69-87.
- MORIN, Y.-C., 1987. « Remarques sur l'organisation de la flexion en français », *ITL Review of Applied Linguistics*, 77-78, 13-91.
- PIRRELLI, V. & BATTISTA, M., 2000. « The Paradigmatic Dimension of Stem Allomorphy in Italian Verb Inflection », *Rivista di Linguistica* 12(2), 307-380.
- SANTOS, D. ET ROCHA, P., 2001. « Evaluating CETEMPúblico, a free resource for Portuguese », in *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, 442-449.
- SIMS, A., 2010. « Probabilistic paradigmatics: Principal parts, predictability and (other) possible particular pieces of the puzzle », présentation au Fourteenth International Morphology Meeting, Budapest.
- SPENCER, A., 2012. « Identifying Stems », *Word Structure*, 5, 88-108.
- STUMP, G. & FINKEL, R., 2013. *Morphological typology: From word to paradigm*, Cambridge, Cambridge University Press.
- STUMP, Gregory. 2001. *Inflectional morphology*. Cambridge : Cambridge University Press.
- THORNTON, A., 2012. « Reduction and maintenance of overabundance. A case study on Italian verb paradigms », *Word Structure* 5, 183-207.
- VEIGA, A., CANDEIAS, S. ET PERDIGÃO, F., 2012. « Generating a pronunciation dictionary for Portuguese using a joint-sequence model with embedded stress assignment », *Journal of the Brazilian Computer Society*, 88.
- VERKUYL, H., VET, C., BORILLO, A., BRAS, M., LE DRAOULEC, A., MOLENDIJK, A., DE SWART, H., VETTERS, C. & VIEU, L., 2004. « Tense and aspect in sentences », in F. Corblin, H. de Swart (eds.), *Handbook of French Semantics*, Stanford, CSLI Publications, 233-270. Stanford : CSLI Publications.
- WURZEL, W. U., 1984. *Flexionsmorphologie und Natürlichkeit. Ein Beitrag zur morphologischen Theoriebildung*, Berlin, Akademie-Verlag.
- ZWICKY, A. 1985. « How to describe inflection », in *Proceedings of the Berkeley Linguistics Society*, vol. 11, 372-386.

Olivier Bonami  
 Université Paris-Sorbonne  
 Institut Universitaire de France  
 Laboratoire de linguistique formelle  
 olivier.bonami@paris-sorbonne.fr

Ana R. Luís  
 Universidade de Coimbra  
 CELGA  
 aluis@fl.uc.pt