



**HAL**  
open science

## Analyse automatique d'espaces thématiques

Gilles Boyé, Anna Kupść

► **To cite this version:**

Gilles Boyé, Anna Kupść. Analyse automatique d'espaces thématiques. 21ème Conférence du Traitement Automatique des Langues Naturelles (TALN), Jun 2014, Marseille, France. hal-01929806

**HAL Id: hal-01929806**

**<https://hal.science/hal-01929806>**

Submitted on 5 May 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Analyse automatique d'espaces thématiques

Gilles Boyé Anna Kupś

Université Bordeaux-Montaigne, Domaine universitaire, 33607 Pessac Cedex  
CLLE-ERSS, UMR5263, CNRS, 5 allées Antonio Machado, 31058 Toulouse Cedex 9  
gboye@u-bordeaux-montaigne.fr, akupsc@u-bordeaux-montaigne.fr

**Résumé.** Basé sur les calculs d'entropie conditionnelle de (Bonami & Boyé, à paraître), nous proposons un analyseur automatique de la flexion dans le cadre de la morphologie thématique qui produit le graphe de régularités du paradigme. Le traitement se base sur un lexique de 6440 verbes extraits du BDLex (de Calmès & Pérennou, 1998) associés à leurs fréquences dans Lexique3 (New *et al.*, 2001). L'algorithme se compose de trois éléments : calcul de l'entropie conditionnelle entre paires de formes fléchies, distillation des paradigmes, construction du graphe de régularités. Pour l'entropie, nous utilisons deux modes de calcul différents, l'un se base sur la distribution de l'effectif des verbes entre leurs différentes options, l'autre sur la distribution des lexèmes verbaux en fonction de leurs fréquences pour contrebalancer l'influence des verbes ultra-fréquents sur les calculs.

**Abstract.** Based on the entropy calculations of (Bonami & Boyé, à paraître), we propose an automatic analysis of inflection couched in the stem spaces framework. Our treatment is based on a lexicon of 6440 verbs present in BDLex (de Calmès & Pérennou, 1998) and associated with their frequencies from Lexique3 (New *et al.*, 2001). The algorithm we propose consists in three steps : computing conditional entropy between all pairs of inflected forms, distilling the paradigms and constructing a regularity graph. For computing entropy, we use two methods : the first one is based on count of verbs in a given distribution whereas the second one takes into account the frequency of each verbal lemma in the distribution to compensate for the bias introduced by the ultra-frequent verbs in the calculation.

**Mots-clés :** morphologie flexionnelle, espaces thématiques, graphe des régularité, français, verbes.

**Keywords:** Inflectional morphology, stem spaces, regularity graph, French, verbs.

## 1 Introduction

Notre travail se situe dans le cadre de la représentation de la flexion en morphologie théorique. Il s'agit ici de présenter un analyseur automatique de la flexion qui produit une représentation des interprédicibilités entre les formes d'un paradigme. Cette représentation dans le cadre de la morphologie thématique prend la forme d'un graphe de régularités (voir Boyé, 2011, pour un exemple de ce type de graphe) qui explicite les liens prédictibles entre les formes.

Notre outil se base sur un lexique de 6440 lexèmes verbaux associés à leurs fréquences. Grâce à BDLex (de Calmès & Pérennou, 1998), une base lexicale de formes fléchies contenant une transcription phonologique de chaque forme, nous constituons le tableau des paradigmes (48 formes par verbe) pour l'ensemble des verbes en API. Les fréquences de lemmes de Lexique3 (New *et al.*, 2001) sont utilisées dans une des deux versions du calcul de l'entropie pour favoriser les généralisations concernant les lexèmes moins fréquents. Ces résultats de prédictibilité basés sur la fréquence donnent une représentation qui pourrait être exploitée pour des études psycholinguistiques dans la lignée de (?).

Notre algorithme se compose de trois éléments : le calcul de l'entropie conditionnelle entre paires de formes (une élaboration sur le calcul de Bonami & Boyé, à paraître), la distillation des paradigmes<sup>1</sup>, construction du graphe de régularités.

---

1. Extraction d'un sous-paradigme suffisant pour capter toutes les régularités/irrégularités.

## 2 Approche thématique

Nous nous situons dans une perspective d'analyse de la flexion au travers des relations entre les formes fléchies. Nous nous basons en partie sur les analyses paradigmatiques développées au sein de l'approche thématique (terme introduit par Plénat, 2008) pour le courant qui considère qu'un lexème est représenté dans le lexique par une collection de radicaux indexés, appelé *espace thématique* (p.ex. Bonami & Boyé, 2003; Boyé & Cabredo Hofherr, 2006; Bonami *et al.*, 2009; Roché, 2010; Tribout, 2012).

Notre analyse morphologique appartient aux approches abstraitives de (Blevins, 2006) qui relie directement les formes de surfaces entre elles sans hypothèses sur des formes sous-jacentes, plutôt qu'aux approches qu'il appelle constructives, caractérisées par des formes sous-jacentes et des opérations de construction conduisant aux formes de surface. Le cadre paradigmatique dans lequel nous nous situons se distingue des théories syntagmatiques décrites par (Stump, 2001) avec ses différents modèles d'association entre une forme et ses morphèmes/exposants.

L'approche thématique en morphologie flexionnelle se concentre depuis ses débuts sur les relations entre les formes fléchies au sein des paradigmes mais en se situant au niveau des allomorphies radicales. La question centrale est celle de la correspondance entre les thèmes indexés, autrement dit le remplissage de l'espace thématique.

Nous reprenons ici l'essentiel de cette approche mais en nous situant à la suite de (Bonami & Boyé, à paraître) au niveau des formes fléchies complètes. Nous utilisons, en particulier, l'interprédictibilité des formes fléchies et les notions de lexèmes réguliers et de sous-régularités pour produire un graphe représentant ces informations.

### 2.1 Interprédictibilité

Dans l'approche thématique, les cases du paradigme dont les contenus sont systématiquement interprédictibles sont formellement associées à un même thème. L'exemple le plus connu en français est celui des formes du futur et du conditionnel qui sont toujours basées sur le même radical quel que soit le verbe considéré. L'ensemble des thèmes nécessaires à la description de la flexion d'une catégorie morphosyntaxique constitue son espace thématique. L'analyse de (Bonami & Boyé, 2003) utilise un espace thématique à 12 cases (thèmes) pour la description de la conjugaison du français. Pour un verbe donné, chaque thème contient un radical qui permet de dériver les formes qui lui sont associées.

	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	T <sub>6</sub>	T <sub>7</sub>	T <sub>8</sub>	T <sub>9</sub>	T <sub>10</sub>	T <sub>11</sub>	T <sub>12</sub>
PASSER	pas	pas	pas	pas	pas	pas	pas	pas	pase	pasə	pasa	pase
BOIRE	byv	bwav	bwa	byv	bwa	byv	bwav	byv	bwa	bwa	by	by
ALLER	al	võ	va	al	va	al	aj	al	ale	i	ala	ale

TABLE 1 – Exemples de radicaux correspondant aux 12 thèmes.

(Bonami & Boyé, à paraître) défendent une approche abstraite qui abandonne l'hypothèse du thème partagé pour adopter la notion des alliances de formes de (Stump & Finkel, 2013). La notion d'alliance permet d'associer les cases du paradigme dont les contenus sont systématiquement interprédictibles directement entre elles sans supposer de forme sous-jacente commune.

	1sg	2sg	3sg	1pl	2pl	3pl
PRÉSENT	3	3	3	1	1	2
IMPARFAIT	1	1	1	1	1	1
PASSÉ	11	11	11	11	11	11
FUTUR	10	10	10	10	10	10
SUBJ. PRÉS.	7	7	7	8	8	7
SUBJ. IMPARF.	11	11	11	11	11	11
CONDITIONNEL	10	10	10	10	10	10
IMPÉRATIF	—	5	—	6	6	—

INFINITIF	9
PART. PRÉSENT	4
PART. PASSÉ	12

TABLE 2 – Les alliances de formes de la conjugaison du français d'après (Bonami & Boyé, 2003)<sup>2</sup>

2. Les cases qui portent le même numéro font partie de la même alliance.

Par exemple, dans le cas des alliances de la Table 2, l'ensemble des cases {imparfait.1sg, présent.3pl, présent.1sg, part.présent, impératif.2sg, impératif.1pl, subj.prés.1sg, subj.prés.1pl, infinitif, futur.1sg, passé.1sg, part.passé}<sup>3</sup> constitue une distillation du paradigme.

Dans ce cadre, on a l'équivalent de l'espace thématique en conservant une forme représentative pour chaque alliance (parties principales). On obtient ainsi un ensemble réduit de formes, distillation du paradigme (terme introduit par Stump & Finkel, 2013), qui capte l'ensemble des données suffisantes pour prédire systématiquement les paradigmes complets.

## 2.2 Régularités et sous-régularités

Dans le cas de l'espace thématique comme avec la distillation du paradigme, l'étape suivante consiste à étudier les relations entre les éléments. Dans la suite de cette section, la description repose sur une distillation et ses parties principales. Dans le cadre initial de l'approche thématique, il s'agissait d'observer les relations entre les formes des lexèmes réguliers et de les comparer avec celles des lexèmes dits irréguliers pour identifier les similarités locales (Bonami & Boyé, 2003, est un exemple typique de ce type d'analyse). Par exemple, bien que MORDRE soit irrégulier, le rapport entre *nous mordons* et *ils mordent* est le même que entre *nous bordons* et *ils bordent* pour BORDER. Dans l'approche thématique, l'apparition de ce type de sous-régularités pour un effectif important de lexèmes irréguliers est associée à une notion de proximité entre les formes. Ces rapprochements sont ensuite utilisés pour établir un graphe reliant les éléments de la distillation entre eux.

Autrement dit, dans le cas de la conjugaison du français, pour chaque type de verbes réguliers, les rapports qu'entretiennent ses parties principales entre elles définissent une classe flexionnelle au sens classique<sup>4</sup>.

	IPF.1	PRS.6	PRS.1	PCP.PRS	IMP.2	IMP.4	SBJV.1	SBJV.4	INF	FUT.1	PST.1	PCP.PST
Régulier.1	X	X	X	X	X	X	X	X	Xe	Xə	Xa	Xe
Régulier.2	Xis	Xis	Xi	Xis	Xi	Xis	Xis	Xis	Xi	Xi	Xi	Xi
SORTIR	sɔʁt	sɔʁt	sɔʁ	sɔʁt	sɔʁ	sɔʁt	sɔʁt	sɔʁt	sɔʁti	sɔʁti	sɔʁti	sɔʁti
MORDRE	mɔʁd	mɔʁd	mɔʁ	mɔʁd	mɔʁ	mɔʁd	mɔʁd	mɔʁd	mɔʁd	mɔʁd	mɔʁdi	mɔʁdy

TABLE 3 – Exemples de sous-régularités.

La comparaison de ces classes flexionnelles régulières avec les paradigmes des verbes irréguliers permet d'identifier les formes correspondant exactement aux mêmes rapports et d'établir toutes les sous-régularités. L'organisation de ces sous-régularités est alors captée sous la forme d'un graphe dont les arcs représentent les sous-régularités. Dans le cas de la Table 3, on observe une sous-régularité qui ne concerne que PRS.1 et IMP.2, on relie donc les deux cases directement par un arc. Pour les sous-régularités concernant des séries plus grandes comme IPF.1, PRS.6, PCP.PRS, IMP.4, SBJV.1, SBJV.4, ci-dessus, on observe les sous-régularités pour l'ensemble des lexèmes et on procède par recouvrements pour inférer les arcs. Jusqu'à présent ce type de graphe a été construit manuellement, nous proposons ici une méthode pour le construire automatiquement à partir d'une approche entropique.

## 3 Approche entropique

(Ackerman *et al.*, 2009) amènent formellement une nouvelle perspective pour la morphologie flexionnelle avec le Paradigm Cell Filling Problem (une version plus générale de la question du remplissage des espaces thématiques) et l'étude des relations entre formes en terme de théorie de l'information (Shannon, 1948). Cette nouvelle approche, basée sur la notion d'entropie conditionnelle, mesure notamment l'incertitude pour le remplissage d'une case du paradigme à partir d'une autre pour l'ensemble des lexèmes. L'entropie est mesurée en bits d'information nécessaires à éliminer l'incertitude.

$$H[X] = - \sum_{x \in X} p(x) \log_2 p(x) \quad (1)$$

Une entropie nulle représente donc l'absence totale d'incertitude, un tirage à pile ou face présente une entropie de 1 et un dé à 6 faces, une entropie de 2,58.

3. La distillation proposée ici est constituée des formes dont les numéros figurent en italique dans la Table 2. Il s'agit de choisir une case représentant chaque alliance, autrement dit une case pour chaque numéro de 1 à 12.

4. Pour des raisons de place, les noms des cases du paradigmes sont abrégés dans la suite : IPF=indicatif imparfait, PRS=indicatif présent, PCP=participe, IMP=impératif présent, SBJV=subjonctif présent, INF=infinitif, PST=passé, 1=1sg, 2=2sg, 3=3sg, 4=1pl, 5=2pl, 6=3pl.

(Bonami *et al.*, 2011), à la suite de (Ackerman *et al.*, 2009), introduisent l’approche entropique pour comparer la complexité du remplissage des paradigmes entre le français et le mauricien. Leur analyse construite autour du moteur de calcul d’entropie PredSPE développé par Olivier Bonami. Le calcul de l’entropie dans notre travail est basé sur une adaptation de ce moteur.

### 3.1 Entropie conditionnelle

Pour calculer l’entropie conditionnelle, nous utilisons le mécanisme proposé par (Bonami & Boyé, à paraître). La procédure se décompose en deux étapes. La première est basée sur le Minimal Generalization Learner de (Albright, 2002) appliquée à chaque paire de cases du paradigme :

- pour chaque lexème, on calcule la transformation phonologique qui permet de passer de la forme<sub>1</sub> à la forme<sub>2</sub> par exemple, pour PRS.6 → INF :
  - $\epsilon n \rightarrow \tilde{u}dr$  (PRENDRE, ...)
  - $\emptyset \rightarrow e$  (PASSER, ...)
- pour chaque transformation, on calcule la généralisation phonologique minimale du contexte d’application qui couvre l’ensemble des lexèmes concernés d’après la décomposition des phonèmes en traits distinctifs. La transformation  $\epsilon n \rightarrow \tilde{u}dr$ , par exemple, ne s’applique qu’à PRENDRE et ses dérivés par préfixation (APPRENDRE, COMPRENDRE, REPRENDRE, etc.). La généralisation minimale doit couvrir toute forme se terminant  $pr\epsilon n$ , c’est à dire que le contexte lui-même doit être :  $*pr-$
- la transformation et le contexte donne une règle de correspondance entre la forme<sub>1</sub> et la forme<sub>2</sub>
  - $\epsilon n \rightarrow \tilde{u}dr / *pr-$
  - $\emptyset \rightarrow e / *[p,t,k,b,d,g,f,s,j,v,z,\zeta,m,n,\eta,j,l,r,w,u,i,y,\epsilon,e,\emptyset,a,u,o,\zeta,\tilde{\epsilon},\tilde{\alpha},\tilde{\omega}]-$

La deuxième étape consiste à associer chaque forme<sub>1</sub> avec l’ensemble des règles de correspondance dont le contexte d’application est compatible. Ces règles forment un n-uplet caractéristique qui à son tour forme une classe de formes<sub>1</sub>. On évalue l’entropie locale pour chaque classe, et à partir de ces entropies locales, on obtient l’entropie conditionnelle de la case<sub>1</sub> à la case<sub>2</sub>.

Par exemple, toujours pour PRS.6 → INF, on a 421 verbes qui sont compatibles avec exactement les deux règles suivantes. Ils forment la classe 7 et leur effectif se répartit comme suit :

417 :  $\emptyset \rightarrow e / *[p,t,k,b,d,g,f,s,j,v,z,\zeta,m,n,\eta,j,l,r,w,u,i,y,\epsilon,e,\emptyset,a,u,o,\zeta,\tilde{\epsilon},\tilde{\alpha},\tilde{\omega}]-$  (ALITER)

4 :  $\emptyset \rightarrow r / *[m,n,\eta,j,l,r,w,u,i,y,\epsilon,e,\emptyset,a,u,o,\zeta,\tilde{\epsilon},\tilde{\alpha},\tilde{\omega}][p,t,k,b,d,g,f,s,j,v,z,\zeta,j,r,w,u,i,y,\epsilon,e,\emptyset,a,u,o,\zeta]-$  (INCLURE)

Cette répartition des effectifs (417 vs 4) donne une entropie conditionnelle locale de 0,0775 qui se combine avec celles de toutes les autres classes de PRS.6 à INF pour donner une entropie conditionnelle de 0,288 pour cette paire de cases.

À la fin de cette phase, on obtient un tableau des entropies conditionnelles entre toutes les cases du paradigme. Comme le font remarquer (Bonami & Boyé, à paraître), ce tableau n’est pas symétrique. Par exemple, la case IPF.1 permet de prédire à coup sûr la case IPF.4 tandis que, en sens inverse, l’entropie est très élevée.

### 3.2 Analyse automatique

Notre analyse automatique, proposée ici, consiste à explorer le tableau des entropies afin de trouver un graphe, qui permette de connecter toutes les cases du paradigme en minimisant les incertitudes, autrement dit, en optimisant l’entropie.

Dans un premier temps, nous procédons à la détection des alliances de formes, c-à-d des formes qui sont interprétables (i.e. prédictibles dans les deux sens). En terme d’entropie, cela signifie qu’il n’y a aucune incertitude pour prédire la case<sub>1</sub> par rapport à la case<sub>2</sub> et inversement. La procédure consiste donc à repérer dans le tableau les paires de cases dont l’entropie conditionnelle est réciproquement nulle. Effectivement, cette méthode nous permet de retrouver, entre autres, l’alliance entre les formes du futur et du conditionnel discutée à la section 2.1.

Ensuite, on procède à la distillation du paradigme. Une fois les alliances identifiées, toutes les formes d’une alliance étant *équivalentes*, on choisit une seule forme comme représentant de chaque alliance. Par exemple, si on garde dans le tableau la forme FUT.1 comme représentant de l’alliance des formes du futur et du conditionnel, toutes les lignes et colonnes qui correspondent aux autres membres de la même alliance, par exemple, FUT.2 ou COND.1, disparaissent.

On obtient ainsi un tableau réduit du même type que le tableau complet mais qui ne contient qu’un représentant pour les formes interprétables. On a alors conservé uniquement les formes qui sont indispensables pour (re-)construire le paradigme complet.

Nous cherchons ensuite à identifier des liens entre les cases retenues, afin de construire le graphe des régularités qui nous permettra de connecter toutes les cases du paradigme.

Pour constituer ce graphe, on explore le tableau distillé. Chaque case correspond à un nœud dans le graphe. Le but est de joindre tous les nœuds, en ajoutant les liens (arcs dans le graphe) les plus sûrs. On commence par mettre en place les relations les plus prédictibles c-à-d par identifier les paires de cases dont les entropies conditionnelles sont les plus faibles. Une fois tous les liens d'un niveau (même valeur d'entropie) exploités, on passe au niveau supérieur pour continuer à ajouter des arcs. Un nouvel arc est ajouté si il permet d'atteindre de nouveaux nœuds. Par exemple, si la case<sub>1</sub> et la case<sub>2</sub> sont déjà connectées, ainsi que la case<sub>2</sub> et la case<sub>3</sub>, on n'ajoutera pas de nouvel arc pour connecter directement la case<sub>1</sub> à la case<sub>3</sub>. La procédure s'arrête quand tous les nœuds peuvent s'inter-atteindre et qu'on a obtenu un graphe orienté connecté.

## 4 Analyse des résultats

À partir du tableau des entropies conditionnelles obtenu grâce à PredSPE, notre analyse automatique a tracé un premier graphe de régularités (à gauche, Figure 1) basé sur la répartition des effectifs de verbes dans les différentes classes. Parmi les 15 alliances distinguées, 6 reposent sur quelques lexèmes très irréguliers qui jouent un rôle majeur dans le calcul des entropies. Ces lexèmes sont tous très fréquents et leurs formes sont pour la plupart mémorisées. Ils se situent en dehors du système de régularités que nous cherchons à identifier ici. Pour minimiser leur influence, nous avons introduit la fréquence comme un des paramètres du calcul et modifié PredSPE en conséquence. Chaque verbe a été associé à l'inverse de sa fréquence de lemme dans Lexique.org<sup>5</sup>. Les verbes dérivés par préfixation ont été regroupés et associés à la fréquence la plus haute de la famille pour tenir compte de l'effet d'entraînement du modèle sur les autres membres de sa famille. Par exemple, *PRENDRE* dont la flexion suit directement le modèle de *PRENDRE* est associé non pas à sa propre fréquence (126/Mmots) mais à celle de son modèle plus fréquent (1913/Mmots). Ce calcul permet de favoriser les généralisations sur les lexèmes à fréquence réduite et donc augmenter l'importance des régularités recherchées. Le graphe des régularités correspondant à ce calcul est à droite dans la Figure 1.

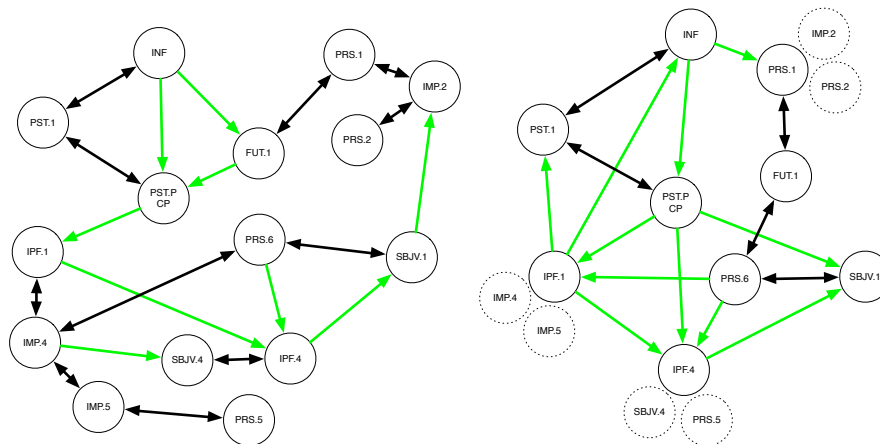


FIGURE 1 – Graphes des régularités basés sur l'entropie avec les effectifs et avec les fréquences<sup>6</sup>

Dans les deux cas, les graphes prennent une forme relativement différente de celles des graphes obtenus manuellement mais on y retrouve des motifs communs. Par exemple, les relations entre INF, PST.PCP et PST.1 ou celle entre FUT.1 et PRS.1 apparaissent aussi bien dans nos deux versions que dans dans le graphe obtenu manuellement (Boyé, 2011). L'introduction des fréquences permet de réduire la distillation de 15 à 9 cases en formant des alliances plus étendues. La réduction de l'influence de verbes extrêmement fréquents, comme *ALLER*, *ÊTRE*, *AVOIR*, *SAVOIR*, *FAIRE* et *DIRE*, a permis de regrouper des alliances du premier graphe dans le second (IPF.1 avec IMP.4 et IMP.5, IPF.4 avec SBJV.4 et PRS.5, PRS.1 avec PRS.2 et IMP.2).

5. Ici, nous avons utilisé la fréquence du lemme dans FranText.

6. Les arcs notés en noir correspondent à des relations d'interprédictibilité symétrique. C'est à dire que la case A et la case B sont les meilleurs prédicteurs l'une de l'autre. Les arcs notés en vert correspondent à des relations asymétriques. C'est à dire que la case A est le meilleur prédicteur de la case B, mais pas l'inverse.

Selon nous, les différences entre les graphes obtenus manuellement et les nôtres tiennent à deux facteurs. D'une part, les sous-régularités recherchées sont différentes. Dans les graphes manuels, on présuppose des modèles réguliers *arbitrairement* et on cherche des analogies similaires pour les irréguliers. Dans notre travail, nous avons cherché à identifier des sous-régularités *sans hypothèse a priori sur des classes régulières*, en étudiant directement toutes les analogies. D'autre part, les relations de nos graphes sont orientées pour tenir compte du fait que la prédictibilité n'est pas symétrique dans le cas général tandis que les relations établies à la main dans le cadre des études thématiques précédentes étaient toutes symétriques par construction, à l'exception de (Bonami & Boyé, 2007) qui, du fait, limitait son étude aux seuls verbes réguliers.

Pour conclure, la méthodologie présentée ici permet donc d'étendre les possibilités de l'approche thématique à des relations plus complexes sur de grands jeux de données. Elle peut s'appliquer à d'autres langues alphabétiques à morphologie affixe. L'outil utilisé peut s'appliquer directement à un lexique flexionnel phonétisable avec, si possible, une base de fréquence pour les lexèmes.

Pour compléter le développement de notre analyseur, il serait nécessaire d'associer automatiquement à chaque arc, les règles de transformation régulières permettant de prédire les correspondances entre formes dans le passage d'une case à une autre. Ceci constituera l'étape suivante de notre travail.

## Références

- 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 : Form and Acquisition*, p. 54–82, Oxford : Oxford University Press.
- ALBRIGHT A. (2002a). *The identification of bases in morphological paradigms*. PhD thesis, UCLA.
- ALBRIGHT A. (2002b). Islands of reliability for regular morphology : evidence from Italian. *Language*, **78**, 684–709.
- BLEVINS J. P. (2006). Word-based morphology. *Journal of Linguistics*, **42**, 531–573.
- BONAMI O. & BOYÉ G. (2003). Supplétion et classes flexionnelles dans la conjugaison du français. *Langages*, **152**, 102–126.
- BONAMI O. & BOYÉ G. (2007). Remarques sur les bases de la conjugaison. In E. DELAIS-ROUSSARIE & L. LABRUNE, Eds., *Des sons et des sens : données et modèles en phonologie et en morphologie*, p. 77–90.
- BONAMI O. & BOYÉ G. (à paraître). De formes en thèmes. In F. VILLOING & S. DAVID, Eds., *Foisonnements morphologiques. Etudes en hommage à Françoise Kerleroux* : Presses Universitaires de Vincennes.
- BONAMI O., BOYÉ G. & HENRI F. (2011). Measuring inflectional complexity : French and Mauritian. Présentation au colloque Quantitative Measures in Morphology and Morphological Development, San Diego, 2011.
- BONAMI O., BOYÉ G. & KERLEROUX F. (2009). L'allomorphie radicale et la relation flexion-construction. In B. FRADIN, F. KERLEROUX & M. PLÉNAT, Eds., *Aperçus de morphologie du français*, p. 267–286.
- BOYÉ 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*, Langues et Syntaxe, p. 41–68 : Hermes Science Publishing.
- BOYÉ G. & CABREDO HOFHERR P. (2006). The structure of allomorphy in Spanish verbal inflection. *Cuadernos de Lingüística*, **13**, 9–24.
- DE CALMÈS M. & PÉRENNOU G. (1998). Bdlex : a lexicon for spoken and written French. In *1st International Conference on Language Resources and Evaluation*, p. 1129–1136. Grenade : ELRA.
- NEW B., PALLIER C., FERRAND L. & MATOS R. (2001). Une base de données lexicales du français contemporain sur internet : Lexique. *L'Année Psychologique*, **101**, 447–462.
- PLÉNAT M. (2008). Le thème l de l'adjectif et du nom. *Congrès Mondial de Linguistique Française 2008*, p. 139.
- ROCHÉ M. (2010). Base, thème, radical. *Recherches linguistiques de Vincennes*, **39**.
- SHANNON C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–423, 623–656.
- STUMP G. T. (2001). *Inflectional Morphology*. Cambridge University Press.
- STUMP G. T. & FINKEL R. (2013). *Morphological Typology : From Word to Paradigm*, volume 138 of *Cambridge Studies in Linguistics*. Cambridge University Press.
- TRIBOUT D. (2012). Verbal stem space and verb to noun conversion in French. *Word Structure*, **5**(1), 109–128.