



Améliorer un lexique syntaxique à l'aide des tables du Lexique-Grammaire : Adverbes en -ment

Benoît Sagot, Karen Fort

► To cite this version:

Benoît Sagot, Karen Fort. Améliorer un lexique syntaxique à l'aide des tables du Lexique-Grammaire : Adverbes en -ment. 26e Colloque International sur le Lexique et la grammaire 2007, Oct 2007, Bonifacio, France. inria-00186779

HAL Id: inria-00186779

<https://inria.hal.science/inria-00186779>

Submitted on 12 Nov 2007

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.

Améliorer un lexique syntaxique à l'aide des tables du Lexique-Grammaire : *Adverbes en -ment*

Benoît Sagot¹, Karën Fort²
INRIA

Abstract

Le *Lefff* (Lexique des Formes Fléchies du Français) est un lexique morphologique et syntaxique du français à large couverture destiné au Traitement Automatique des Langues (TAL), en particulier pour l'analyse syntaxique. C'est sur les verbes pleins que se sont concentrés la plupart des travaux récents sur l'enrichissement du *Lefff*. Toutefois, nous montrons ici comment nous avons enrichi le *Lefff* à partir des tables du Lexique-Grammaire des adverbes en *-ment*, publiées par Molinier (Molinier & Levrier 2000) et intégralement disponibles au format électronique sous licence libre. Ceci a nécessité un double travail de linguistique et de modélisation, afin d'exploiter leur contenu dans un lexique TAL tel que le *Lefff*. Nous terminons par une brève évaluation du lexique obtenu, en nous restreignant, faute de points de comparaison au niveau syntaxiques, à sa dimension morphologique.

1. Introduction

L'analyse syntaxique profonde ne peut se faire qu'à la condition de disposer d'informations lexicales riches qui forment un lexique morphologique et syntaxique du français à large couverture, pour permettre à la fois la *construction* d'analyses linguistiquement riches, et la *désambiguïsation*, grâce à des contraintes fines. De plus, un tel lexique doit représenter ces informations d'une façon adaptée à l'utilisation dans les systèmes de traitement automatique. Une telle ressource n'est pas encore disponible pour le français, malgré de nombreuses initiatives indépendantes, qui ont atteint des degrés d'avancement divers et qui ont privilégié des aspects différents. Deux d'entre elles ont donné naissance à des ressources à large couverture que tout sépare :

- les tables du Lexique-Grammaire, dont le développement a été initié par Maurice Gross au LADL (Gross 1975) et se poursuit à l'IGM autour d'Éric Laporte ;
- le *Lefff* (Lexique des Formes Fléchies du Français), lexique morphologique et syntaxique du français à large couverture (520 000 entrées) utilisé dans divers systèmes de TAL (Sagot *et al.* 2006; Sagot & Danlos 2007).

Des travaux visant à exploiter les tables du Lexique-Grammaire pour l'enrichissement et la correction du *Lefff* ou d'autres lexiques syntaxiques ont déjà été publiés (Sagot & Danlos 2007; Danlos *et al.* 2006; Gardent *et al.* 2005). Toutefois, ces travaux se concentrent sur les entrées verbales et leurs cadres de sous-catégorisation. Si un tel effort est indispensable et sera poursuivi, il n'en reste pas moins que les autres catégories sont également à prendre en compte.

¹ INRIA Futurs - projet Signes, Dom. Universitaire, 351 cours de la Libération, 33405 Talence Cedex, France, benoit.sagot@inria.fr

² Loria/INRIA Lorraine - projets TALARIS et Calligramme, Campus Scientifique - UMR 7503 - 54506 Vandœuvre-lès-Nancy, France, karen.fort@loria.fr

Parmi ces autres catégories, il en est une dont la variabilité interne est très importante mais quasiment ignorée des lexiques syntaxiques utilisés en TAL comme le *Lefff* : les adverbes. De fait, un traitement quasi uniforme de l'ensemble des adverbes, tel qu'il est pratiqué aujourd'hui dans la plupart des analyseurs, induit une sur-génération très importante, entraînant des pertes de performances en termes d'ambiguïté, de temps d'analyse et de précision. De plus, les analyses produites ne reflètent pas la richesse des modificateurs adverbiaux. Or, dans le cas particulier des adverbes en *-ment*, et contrairement à ce qui se passe pour d'autres catégories, l'intégralité des tables du Lexique-Grammaire est disponible librement au format électronique, sous licence LGPL-LR. Il s'agit des tables dites de Molinier et décrites dans (Molinier & Levrier 2000). Les tables d'adverbes figés de M. Gross (Gross 1986) sont également disponibles. Toutefois, les problèmes spécifiques posés par le figement et le nombre pléthorique d'entrées discutables dans ces tables nous a conduits à nous concentrer pour l'instant, sur les adverbes en *-ment*.

Nous présentons successivement les tables du Lexique-Grammaire pour les adverbes en *-ment*, puis leur exploitation en vue de l'enrichissement du *Lefff*, et enfin une rapide évaluation du lexique obtenu en comparant la liste des entrées avec d'autres lexiques morphologiques. L'absence de point de comparaison quant aux informations syntaxiques ne permet pas, en effet, de proposer une évaluation qui aille au-delà.

2. Tables du Lexique-Grammaire pour les adverbes en *-ment*

2.1. Données lexicales

Comme le détaille (Molinier & Levrier 2000), les adverbes en *-ment* forment une classe numériquement importante d'adverbes, qui, contrairement aux autres adverbes, est une classe ouverte. Ces adverbes forment une classe morphologiquement homogène, puisqu'ils sont majoritairement construits sur un schéma du type adjectif+*ment*. Il existe toutefois quelques exceptions : certains adverbes sont construits sur des interjections (*fichtrement*, *diablement*,...), d'autres sur des adverbes (*quasiment*), d'autres enfin sur des réfections de formes anciennes (ainsi *nuitamment* à partir de *nuitantre*).

Les tables données par (Molinier & Levrier 2000), et dont la version électronique actuelle, disponible librement sous licence LGPL-LR, diffère peu, ont été constituées grâce à l'examen de tous les adverbes en *-ment* présents dans les grands dictionnaires de la langue du XX^e siècle : le Trésor de la Langue Française, le Grand Larousse de la langue française et le Grand Robert de la langue française. Ont ainsi été recueillis 2330 adverbes en *-ment*, auxquels ont été ajoutés 450 adverbes en *-ment* d'usage courant mais absents de ces ouvrages. Au total, (Molinier & Levrier 2000) comporte donc 2780 formes adverbiales correspondant à 3170 entrées (lignes) dans 9 tables. On notera que les tables disponibles au format électronique sont actuellement au nombre de 16, un découpage plus fin ayant été effectué. Ces tables contiennent en outre un nombre un peu supérieur d'adverbes (2784 formes pour 3202 entrées)¹.

2.2. Classification générale

Dans (Molinier & Levrier 2000) est établie une première partition entre adverbes de phrase ou adverbes de statut périphérique et adverbes intégrés à la proposition, c'est-à-dire rattachés au verbe ou à tout autre constituant de la proposition.

¹ Dans (Molinier & Levrier 2000), des adverbes autres que les adverbes en *-ment* sont indiqués (mais pas dans les tables en fin d'ouvrage, ni bien sûr dans les tables électroniques). Ces adverbes, souvent des adverbes composés voire des locutions adverbiales productives, posent des problèmes spécifiques : propriétés particulières, problèmes liés au figement et au semi-figement, etc. Leur étude dépasse le cadre du travail présenté ici.

Les *adverbes de phrase* sont identifiables grâce aux deux propriétés suivantes :

1. Adv, N0 ne V pas W = + : possibilité de figurer en position détachée en tête de phrase négative (cf. exemple 1a)
2. C'est Adv que P = - : impossibilité d'extraction dans *C'est...que* (cf. exemple 1b)

La première propriété permet d'établir la portée de l'adverbe sur l'ensemble de la phrase alors que la seconde permet de vérifier l'absence de tout rapport au verbe. Ainsi, *sincèrement* est considéré comme un adverbe de phrase, car il vérifie ces deux propriétés :

- 1a. *Sincèrement, je n'ai pas bien vu*
- 1b. **C'est sincèrement que j'ai bien vu*

des adverbes de phrase et vérifient la disjonction de la négation des propriétés précédemment définies pour les adverbes de phrase. Ces adverbes doivent donc vérifier l'une et/ou l'autre des propriétés suivantes :

1. Adv, N0 ne V pas W = - : impossibilité de figurer en position détachée en tête de phrase négative (cf. exemple 2a)
2. C'est Adv que P = + : possibilité d'extraction dans *C'est...que* (cf. exemple 2b)

Ainsi, *calmement* est considéré comme un adverbe intégré à la proposition, car il vérifie ces deux propriétés :

- 2a. **Calmement, je n'ai pas attendu*
- 2b. *C'est calmement que j'ai attendu*

Les adverbes qui vérifient ces deux propriétés n'ont pas de portée sur la phrase globale et sont des modificateurs du verbe. Ils expriment la manière.

Certains adverbes vérifient la première propriété mais pas la seconde, car ils sont trop étroitement liés au verbe. C'est le cas des adverbes figés avec le verbe (cf. exemples 2c et 2d) ou des adverbes de quantité, d'intensité, de complétude.

- 2c. **Pertinemment, je ne le sais pas*
- 2d. **C'est pertinemment que je le sais*

Enfin, certains adverbes vérifient la seconde propriété, mais pas la première. C'est le cas, en particulier, de certains adverbes de temps et de tous les adverbes de point de vue (cf. exemples 2e et 2f).

- 2e. *Juridiquement, je ne suis pas responsable*
- 2f. *C'est juridiquement que je suis responsable*

suivantes.

2.3. Adverbes de phrase

Les adverbes de phrase sont regroupés dans des tables nommées *P_adv**. Les fichiers correspondants, au nombre de 6, détaillent trois grandes classes d'adverbes de phrase : les conjonctifs, les disjonctifs de style et les disjonctifs d'attitude.

Les conjonctifs (table *P_advpc*), nécessitent un contexte gauche auquel ils renvoient. C'est le cas par exemple dans la phrase *Hâte-toi, autrement tu seras en retard*. Le français compte environ 150 adverbes conjonctifs, dont une cinquantaine sont des adverbes en *-ment*.

Les disjonctifs regroupent les autres adverbes de phrase. Parmi ceux-ci, les disjonctifs de style (table *P_advps*) qualifient la manière de présenter les choses par le locuteur : *Honnêtement, c'est un fiasco*. On compte une soixantaine de disjonctifs de style, dont une quinzaine en *-ment*.

Les autres adverbes disjonctifs sont les disjonctifs d'attitude, qui se subdivisent en 4 sous-classes : une dizaine de disjonctifs d'attitude évaluatifs (table *P_advpa_e*, ex. : *curieusement*), une dizaine de disjonctifs d'attitude d'habitude (table *P_advpa_h*, ex. : *habituellement*), une trentaine de disjonctifs d'attitude modaux (table *P_advpa_m*, ex. : *certainement*) et une trentaine de disjonctifs d'attitude orientés vers le sujet (table *P_advpa_s*, ex. : *sottement*).

La plupart de ces classes sont également définies par la présence ou l'absence de certaines propriétés. Faute de place, nous renvoyons à (Molinier & Levrier 2000) pour plus de détails.

2.4. Adverbes intégrés à la proposition

Les adverbes intégrés à la proposition sont organisés en tables nommées *P_adv_m**, réparties dans 10 fichiers. Ces fichiers détaillent 6 grandes classes d'adverbes intégrés à la proposition : les adverbes de manière orientés vers le sujet, les adverbes de manière verbaux, les adverbes de manière quantifieurs (une table découpée en trois fichiers), les adverbes de point de vue, les adverbes de temps (une table découpée en trois fichiers) et les adverbes focalisateurs.

Les adverbes de manière orientés vers le sujet (table *P_adv_{ms}*) vérifient les propriétés suivantes :

- Adv, N0 ne V pas W = - : inacceptabilité en position détachée en tête de phrase négative,
- C'est Adv que P = + : possibilité d'extraction dans *C'est...que*,
- Nhum être Adj = + : possibilité, pour toute phrase à sujet humain où ils apparaissent, d'association à une phrase prédicative où l'adjectif radical qualifie ce même sujet humain.

Ces adverbes sont près de 900 (table *P_adv_{ms}*). Exemple : *Max regarde anxieusement l'horizon*.

Les adverbes de manière verbaux (table *P_adv_{mv}*) vérifient les propriétés suivantes :

- Adv, N0 ne V pas W = - : inacceptabilité en position détachée en tête de phrase négative,
- C'est Adv que P = *variable* (selon que l'adverbe est libre ou figé),
- Nhum être Adj = - : s'ils apparaissent dans une phrase à sujet humain, elle n'équivaut pas à une phrase prédicative où l'adjectif radical qualifie ce même sujet.

Cette classe est largement ouverte, la table *P_adv_{mv}* contient ainsi plus de 1500 adverbes, grâce à un important travail d'ajout des auteurs (300 formes) en complément des adverbes présents dans les dictionnaires de référence. Exemple : *Max regarde fixement l'horizon*.

Les adverbes de manière quantifieurs (tables *P_adv_{mq}**) vérifient les mêmes propriétés que les adverbes de manière verbaux. Ils offrent en outre la possibilité d'association à un adverbial de quantité ou d'intensité foyer d'une phrase interrogative. Exemple : *Max aime énormément ce tableau*. Les adverbes indiquant la complétude (*partiellement*) ou l'extension qualitative (*fondamentalement*) en font également partie. Ils représentent respectivement 13 et 21 adverbes, les autres quantifieurs étant près de 250.

Les adverbes de point de vue (table *P_adv_{mp}*) vérifient quant à eux les propriétés suivantes :

- Adv, N0 ne V pas W = + : acceptabilité en position détachée en tête de phrase négative,
- C'est Adv que P = *variable* (selon que l'adverbe est libre ou figé),
- d'un point de vue Adj = + : possibilité de paraphrase par *d'un point de vue Adj*.

Exemple : *Légalement, je suis responsable*. Ces adverbes foisonnent, en particulier dans les textes scientifiques et techniques. En outre, de tels adverbes peuvent être créés très facilement, à partir d'un adjectif exprimant une notion, une activité ou une connaissance. Cette classe est donc ouverte et comprend près de 280 adverbes, dont 190 proviennent des dictionnaires de référence. Là encore, un important travail de complétion a été effectué par les auteurs.

Les adverbes de temps (tables *P_adv_{mt}**) vérifient les propriétés suivantes :

- Adv, N0 ne V pas W = *variable*,

- C'est Adv que P = + : possibilité d'extraction dans *C'est...que*,
- Possibilité d'association à un adverbial de temps foyer d'une phrase interrogative.

Exemple : *Je suis venue ici dernièrement*. On distingue dans cette classe les adverbes de durée (table *P_advmtd*, 11 entrées), les adverbes de fréquence (table *P_advmtf*, 33 entrées) et les autres adverbes de temps (table *P_advmtq*, 31 entrées).

Enfin, les adverbes focalisateurs (table *P_advmf*) vérifient les propriétés suivantes :

- Adv, N0 ne V pas W = - : impossibilité de la position détachée en tête de phrase négative,
- C'est Adv que P = - mais C'est Adv Ni que P-Ni = + : impossibilité d'extraction dans *C'est...que*, mais possibilité en compagnie d'un groupe nominal majeur.

C'est une classe fermée avec 16 adverbes en *-ment*. Exemple : *Je lis principalement des romans*.

3. Extraction d'un lexique TAL des adverbes en *-ment* au format Lefff

Les tables décrites dans (Molinier & Levrier 2000) sont intégralement disponibles, sous une licence libre (LGPL-LR). Elles peuvent donc être exploitées pour l'amélioration d'un lexique syntaxique du français destiné au TAL tel que le Lefff. Après une brève présentation du Lefff, nous précisons lesquelles des propriétés présentes dans les tables sont pertinentes pour un lexique syntaxique TAL, puis, à l'inverse, quelles propriétés pertinentes ne sont pas présentes dans les tables. Cette section se termine par une description de la conversion proprement dite et du lexique obtenu. Une évaluation succincte de ce lexique fait l'objet de la section suivante.

3.1. Présentation du Lefff et de son architecture

Le Lefff (Lexique des formes fléchies du français) est un lexique syntaxique à large couverture pour le français (Sagot *et al.* 2006; Sagot & Danlos 2007). Actuellement en version 2.5, il est téléchargeable sous une licence libre (LGPL-LR) à l'adresse <http://www.lefff.net>.

Le Lefff repose sur une architecture à deux niveaux :

- un *lexique intensionnel*, qui associe à chaque lemme une classe morphologique et une classe syntaxique ; c'est à ce niveau qu'est fait le travail de développement ;
Exemple : `manger v-er:std @verbe_transitif_direct`
- un *lexique extensionnel*, obtenu à partir du lexique intensionnel par compilation, qui associe à chaque forme une structure représentant explicitement toutes les informations linguistiques associées ; c'est ce lexique qui est utilisé par les analyseurs.

Exemple : `mange v [pred='manger_1<Suj:(snlcln,Obj:(snlcla)>', cat=v, @pers, @PS13s]`.

Au niveau intensionnel, les informations syntaxiques sont donc décrites à l'aide de classes syntaxiques, définies par héritage de propriétés syntaxiques atomiques, propriétés elles-mêmes définies de façon indépendante de la définition des classes. Au niveau extensionnel, le cadre de sous-catégorisation d'une forme donnée est constitué d'une liste de *fonctions syntaxiques*, chacune indiquant les *réalisations* possibles de cette fonction ainsi que le caractère obligatoire ou non de sa réalisation (indiqué par des parenthèses). La structure syntaxique complète, outre le cadre, comporte le cas échéant des couples attributs-valeurs ou fonctions-structures sous-spécifiées, éventuellement factorisées sous forme de *macros* (introduites par « @ »), et des informations syntaxiques complémentaires (contrôle, attribution, (im)personnel, et tout autre trait syntaxique pertinent en vue de l'analyse syntaxique).

Les fonctions syntaxiques sont définies dans le Lefff par des critères initialement développés pour les entrées lexicales utilisables dans des constructions à tête verbale (verbes, adjectifs et noms prédicatifs) reposent sur la substituabilité (en prenant en compte pronoms *et* syntagmes), sur le principe de réalisation unique d'une fonction syntaxique pour un prédicat donné, et sur

l'identification de la fonction par un paradigme de pronoms (à l'exception des cas à partage d'arguments, c'est-à-dire les attributs). Pour les adverbes, et en particulier les adverbes en *-ment*, la situation est à la fois plus simple, en raison du faible nombre de compléments différents que peut sous-catégoriser un adverbe, mais également plus délicate, puisque les critères reposant sur les pronoms clitiques ne sont pas applicables. Toutefois, l'application des autres critères (et en particulier l'utilisation des pronoms pleins) permet de restreindre à deux le nombre de fonctions syntaxiques sous-catégorisables par un adverbe en *-ment* :

- Objà, lorsqu'un adverbe introduit un complément prépositionnel en *à* (*contrairement à lui*),
- Objde, lorsque le complément est en *de* (*indépendamment de lui / de ce que P*).

Dans le *Lefff*, les réalisations possibles, quant à elles, sont de trois types : *pronoms clitiques*, *syntagme direct* (syntagme nominal (sn), adjectival (sa), infinitif (sinf), phrastique fini (scompl), interrogative indirecte (qcompl)) et *syntagme prépositionnel* (syntagme direct précédé d'une préposition, comme *de-sn*, *à-sinf* ou *pour-sa* ; *à-scompl* et *de-scompl* représentent les réalisations en *à/de ce que P*). Pour les adverbes, seules sont concernées les réalisations prépositionnelles, qui peuvent être nominales (*contrairement à Pierre*) ou phrastiques finies (*indépendamment de ce que Pierre est intelligent*).

3.2. Sélection des propriétés pertinentes dans les tables

La version électronique actuellement disponible des tables de (Molinier & Levrier 2000) décrites à la section précédente associe aux différents adverbes en *-ment* recensés un certain nombre de propriétés, que l'on peut regrouper en quatre grandes familles : les *propriétés macrosyntaxiques* (possibilités de détachements, de constructions clivées, ...), les *propriétés locales* (constructions très locales dans lesquelles l'adverbe peut ou ne peut pas prendre place), les *propriétés paradigmatisques* (quel type d'adverbe interrogatif correspond à l'adverbe étudié) et les *propriétés paraphrastiques* (de type Adv, P = Que P être Adj, où Adj désigne l'adjectif sur la base duquel l'adverbe Adv est construit).

Parmi ces propriétés, nous avons choisi de ne conserver que celles qui ont un intérêt direct pour les analyseurs syntaxiques actuels. Naturellement, ceci ne signifie pas que les propriétés écartées sont sans intérêt, bien au contraire. Ainsi, les propriétés paraphrastiques ont été écartées. En effet, il n'est pas du ressort d'un analyseur syntaxique de savoir que pour tel adverbe mais pas pour tel autre, une construction de type Adv, P est en relation de paraphrase avec une construction de type N0 être Adj de V-inf W (où naturellement Adj est l'adjectif sur lequel Adv est construit, et P = N0 V W). Les propriétés paradigmatisques, qui sont de la forme Adv = Comment, À quelle fréquence, À quel moment, En combien de temps², ont été conservées, car elles permettent de déterminer qu'un adverbe est susceptible ou non de saturer certaines fonctions syntaxiques (temporelle, de manière, etc. ...). Les propriétés locales n'ont été conservées que dans la mesure où elles apportent quelque chose à l'analyse syntaxique : soit elles concernent la valence de l'adverbe, soit elles signalent qu'une construction non-standard est possible. Des exemples respectifs de ces deux cas sont les propriétés Adj-ment de N (fonction Objde sous-catégorisée, réalisation nominale) et Adv parlant (possibilité de former une locution adverbiale de type *linguistiquement parlant*). En revanche, le fait que tel ou tel adverbe admette pour intensificateur les adverbes *tout*, *très*, *plus* et/ou *bien* nous semble de nature à complexifier le lexique sans apporter aux analyseurs syntaxiques une information utile pour la désambiguïsation. Enfin, les propriétés macrosyntaxiques ont été conservées, sauf celles que les principaux analyseurs syntaxiques actuels ne sauraient pas exploiter. Ont ainsi été conservées des proprié-

² Bien que dans le cas de *À quelle fréquence* et *À quel moment*, la terminologie employée préfère utiliser respectivement *souvent*, qui n'est pas un interrogatif, et *quand*, qui est peut-être trop ambigu.

tés permettant ou non divers types de constructions clivées, à détachement ou à extraposition, mais pas celles décrivant les particularités des phrases interrogatives et exclamatives.

3.3. Autres propriétés déjà présentes dans le *Lefff*

Naturellement, le *Lefff* incluait déjà un lexique des adverbes en *-ment*. Pour la plupart, aucune information syntaxique n'était donnée. Mais pour certains d'entre eux, des propriétés pertinentes étaient indiquées, propriétés qui ne font *pas* partie des propriétés que l'on peut extraire des tables de (Molinier & Levrier 2000). Il s'agit principalement de l'attribut *adv_kind*, qui peut prendre trois valeurs (en plus de la valeur par défaut qui n'apporte aucune information) :

- *modnc* : entrées pouvant modifier un groupe nominal ou un déterminant (pour l'instant, le *Lefff* est imprécis sur ce point) : cf. *seulement* dans *seulement 3 pommes* ou *seulement des pommes*³ ; en réalité, tous les focalisateurs de (Molinier & Levrier 2000) sont concernées ;
- *modprep* : entrées pouvant modifier une préposition : cf. *exactement dans la cible* ; certains focalisateurs sont concernés, mais pas tous, et pas uniquement eux (cf. *rigoureusement*) ;
- *très* : concerne les entrées qui, comme l'unique entrée pour l'adverbe *très*, servent exclusivement à modifier des adverbes ; toutefois, aucun adverbe en *-ment* ne semble concerné ;

Nous avons donc conservé ces propriétés, en les harmonisant avec celles extraites des tables.

3.4. Conversion et lexique obtenu

La structure en hiérarchie de classes syntaxiques utilisée dans le *Lefff* intensionnel pour décrire les propriétés syntaxiques des entrées peut être vue comme une généralisation du découpage en tables. De plus, les tables ayant été regroupées en familles, nous avons reproduit la même hiérarchie sous forme de hiérarchie de classes syntaxiques au sens du *Lefff*. Dans les cas, assez nombreux, où le comportement au sein d'une table n'était pas homogène (malgré la non-prise en compte de certaines colonnes), nous avons créé de nouvelles classes syntaxiques qui héritent de la classe correspondant à la table. Au total, nous avons obtenu 37 classes syntaxiques.

Le résultat est un lexique syntaxique orienté TAL des adverbes en *-ment* au format *Lefff*. Le niveau intensionnel comporte 2784 entrées, comme indiqué précédemment. Chacune d'entre elles est associée à une classe morphologique triviale (une seule forme, identique au lemme) et à l'une des 37 classes syntaxiques. Le résultat du processus de compilation est un lexique extensionnel de 3202 entrées pour 2784 formes adverbiales.

4. Évaluation et intégration au *Lefff*

Le lexique obtenu est difficile à évaluer au niveau syntaxique, puisqu'aucune autre référence ne semble disponible. La seule façon de l'évaluer serait d'adapter des analyseurs syntaxiques pour qu'ils tirent pleinement parti de ces informations syntaxiques, puis d'étudier l'amélioration en termes de temps d'analyse, d'ambiguïté et de précision que l'on obtient. Ceci reste à faire. Il est toutefois aisé de comparer ce lexique, noté LG-adv-ment, à d'autres lexiques au niveau morphologique. À cet effet, nous avons extrait les adverbes en *-ment* de trois ressources importantes disponibles pour le français :

- le *Lefff* dans sa version 2.5, c'est-à-dire avant le travail présenté dans cet article, noté *Lefff*_{2.5},
- le lexique morphologique MULTEX du français (Veronis 1998), noté MULTEX-fr,
- le lexique morphologique Morphalou (version 2.0) (Romary *et al.* 2004).

³ En particulier, on remarquera que *C'est seulement des pommes que Pierre a le droit de manger* est correct, alors que **C'est des pommes que Pierre a seulement le droit de manger* ne l'est pas. On retrouve la propriété C'est Adv Ni que P-Ni.

Ressource	Total	Présents dans LG-adv-ment	Absents de LG-adv-ment	
			Intégrés au Lefff _{2.6}	Non intégrés au Lefff _{2.6}
LG-adv-ment	2784	(2784)	–	–
Lefff _{2.5}	1357	1339	15	3
MULTEX-fr	712	699	8	5
Morphalou	1407	1363	37	9

TAB. 1. Comparaison des différents lexiques (morphologiques) d'adverbes en -ment.

Enfin, nous avons effectué un travail de fusion pour intégrer à une nouvelle version du Lefff, la version 2.6, outre les 3202 entrées du lexique LG-adv-ment, les entrées correspondant aux adverbes présents dans les autres ressources, manquant dans LG-adv-ment, et dont la présence nous a semblé pertinente (cf. table 1). Parmi les candidats à l'ajout, un certain nombre ont été rejetés, qui étaient presque tous des erreurs (orthographe erronée voir confusion avec un nom commun). Parmi ceux qui ont été ajoutés, la majorité sont des entrées nouvelles, auxquelles nous avons dû attribuer manuellement une classe syntaxique, mais 5 sont des variantes orthographiques d'entrées présentes dans LG-adv-ment. Parmi les adverbes présents dans telle ou telle des trois autres ressources mais pas dans LG-adv-ment, on peut citer *nonchalamment* (dans les trois), *maladivement* (Lefff_{2.5} et Morphalou), *congrûment*, (MULTEX-fr et Morphalou), *floristiquement* (Lefff_{2.5}), *microbiologiquement* (MULTEX-fr), *bourrument* (Morphalou).

5. Conclusion

Nous avons montré comment nous avons tiré parti des tables du Lexique-Grammaire des adverbes en -ment pour enrichir le Lefff, lexique syntaxique pour le TAL. Nous avons montré comment certaines informations apportées par les tables ont été prises en fusionnées à d'autres, bien moins nombreuses, déjà présentes dans le Lefff. Le résultat est un lexique syntaxique des adverbes en -ment. Il est librement accessible au sein du Lefff 2.6.

Outre la poursuite de ce type de travaux sur d'autres catégories, poursuivant les efforts accomplis sur les constructions verbales impersonnelles et entamés sur les expressions verbales figées, nous avons l'intention de comprendre dans quelle mesure ces enrichissements augmentent la qualité et la richesse des analyseurs syntaxiques qui reposent sur le Lefff.

References

- DANLOS L., SAGOT B. et SALMON-ALT S. (2006), « French frozen verbal expressions : from lexicon-grammar to NLP applications », in *Actes du Colloque Lexique Grammaire 2006*, Palerme, Italie.
- GARDENT C., GUILLAUME B., PERRIER G. et FALK I. (2005), « Maurice Gross' Grammar Lexicon and Natural Language Processing », in *Proc. of the 2nd LTC*, Poznań, Poland.
- GROSS M. (1975), *Méthodes en syntaxe*, Hermann, Paris, France.
- GROSS M. (1986), *Grammaire transformationnelle du français - Syntaxe de l'adverbe*, vol. 3, ASSTRIL, Paris, France.
- MOLINIER C. et LEVRIER F. (2000), *Grammaire des adverbes. Description des formes en -ment*, Droz, Genève, Suisse.
- ROMARY L., SALMON-ALT S. et FRANCOPOULO G. (2004), « Standards going concrete : from LMF to Morphalou », in *Actes du Workshop on Electronic Dictionaries de Coling 2004*, Genève, Suisse.
- SAGOT B., CLÉMENT L., VILLEMONT DE LA CLERGERIE E. et BOULLIER P. (2006), « The Lefff 2 syntactic lexicon for French : architecture, acquisition, use », in *Proc. of LREC'06*.
- SAGOT B. et DANLOS L. (2007), « Améliorer un lexique syntaxique à l'aide des tables du lexique-grammaire – Constructions impersonnelles », in *Cahiers du Cental*, to appear.
- VERONIS J. (1998), *Multext-Lexicons, A set of Electronic Lexicons for European Languages*.