

La ressource Nomage. Confronter les attentes théoriques aux observations du comportement linguistique des nominalisations en corpus

Antonio Balvet, Lucie Barque, Marie H el ene Condette, Pauline Haas, Richard
Huyghe, Rafael Marin, Aur elie Merlo

► To cite this version:

Antonio Balvet, Lucie Barque, Marie H el ene Condette, Pauline Haas, Richard Huyghe, et al.. La ressource Nomage. Confronter les attentes th eoriques aux observations du comportement linguistique des nominalisations en corpus. Traitement Automatique des Langues, ATALA, 2011, Ressources Linguistiques Libres, 52 (3), pp.129-152. <halshs-01077819>

HAL Id: halshs-01077819

<https://halshs.archives-ouvertes.fr/halshs-01077819>

Submitted on 27 Oct 2014

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 destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.



La ressource Nomage

Confronter les attentes théoriques aux observations du comportement linguistique des nominalisations en corpus

A. Balvet* — L. Barque** — M.-H. Condette* — P. Haas** —
R. Huyghe*** — R. Marín* — A. Merlo*

* STL, CNRS UMR 8163, Université Lille 3
prenom.nom@univ-lille3.fr
marie-helene.condette@u-picardie.fr
aurelie.merlo@etu.univ-lille3.fr

** LDI, CNRS UMR 7187, Université Paris 13
prenom.nom@ldi.univ-paris13.fr

*** CLILLAC-ARP, EA 3967, Université Paris 7
prenom.nom@eila.univ-paris-diderot.fr

RÉSUMÉ. La ressource présentée dans cet article combine un corpus de noms déverbaux annotés sémantiquement et syntaxiquement, fondé sur le French Treebank, et un lexique électronique fournissant des informations d'ordre morphologique, syntaxique et sémantique sur les noms présents dans le corpus ainsi que sur les verbes dont ils sont dérivés.

ABSTRACT. The resource presented in this paper combines a semantically and syntactically annotated corpus of deverbal nouns based on the French Treebank, and an electronic lexicon, providing descriptions of morphological, syntactic and semantic properties of the deverbal nouns found in our corpus and of their verbal sources.

MOTS-CLÉS : corpus annoté, lexique fondé sur corpus, nominalisation, aspect lexical, structure argumentale.

KEYWORDS: annotated corpus, corpus-driven lexicon, nominalization, lexical aspect, argumental structure.

1. Introduction

La ressource lexicale présentée dans cet article a été développée dans le cadre du projet Nomage¹. Elle combine un corpus de noms déverbaux annotés sémantiquement et syntaxiquement, basé sur le French Treebank (Abeillé *et al.*, 2003), et un lexique fournissant diverses informations linguistiques sur les noms présents dans le corpus. La ressource est consultable et téléchargeable sous la forme d'un fichier XML (norme LMF) sur le site du projet².

La plupart des travaux portant sur les nominalisations verbales (c'est-à-dire les noms morphologiquement dérivés de verbes) sont consacrés à l'étude de leur structure argumentale et thématique, explorant leurs formes de complémentation, leur distribution avec les déterminants, etc. (Lees, 1960; Chomsky, 1970; Milner, 1982; Grimshaw, 1990; Alexiadou, 2001). En complément de ces approches essentiellement syntaxiques, le projet Nomage s'est proposé d'étudier les propriétés sémantiques des nominalisations, et notamment leurs propriétés aspectuelles, telles qu'elles peuvent être héritées des verbes correspondants. L'originalité du travail proposé repose sur l'utilisation et la comparaison de deux méthodes de description appliquées à un même corpus de lexèmes. La première consiste à donner une description de chaque lexème en appliquant une batterie de tests sur des exemples construits. Cette description repose sur une introspection linguistique et a été réalisée par les linguistes membres du projet. La seconde consiste en un étiquetage empirique des occurrences des nominalisations en corpus (en l'occurrence le French TreeBank), étiquetage fondé sur l'application d'une série de tests appliqués à chaque occurrence. La ressource proposée présente ainsi une description des nominalisations issue de ces deux méthodes. Le lexique inclut des informations morphologiques, syntaxiques et sémantiques (notamment aspectuelles) sur les noms étudiés ainsi que sur leurs bases verbales. Le corpus regroupe les occurrences des noms décrits dans le lexique, annotées suivant les résultats aux tests appliqués en contexte³.

Parmi les ressources déjà existantes, le lexique Nomage se rapproche du lexique Verbaction (Hathout *et al.*, 2002) qui liste les couples verbes/noms apparentés, munis de catégories morphosyntaxiques mais sans caractérisation sémantique ni information sur la structure argumentale des unités listées. Le lexique Nomage s'apparente également au lexique Nomlex (Macleod *et al.*, 1998), qui propose une liste de nominalisations en anglais accompagnées de leur structure argumentale, elle-même mise en relation avec les arguments du verbe de base. La ressource développée se rap-

1. Projet « ANR-JC Analyse sémantique et codification lexicale des nominalisations », porté par R. Marín, Université Lille 3, laboratoire STL, 2007-2010.

2. La ressource est téléchargeable à l'adresse <http://sites.google.com/site/nomagesite/delivrables>. L'interface de consultation se trouve quant à elle à l'adresse <http://nomage.recherche.univ-lille3.fr/nomage/>.

3. Les deux méthodes d'analyse mènent ainsi à deux classifications aspectuelles distinctes. D'une part, une classification des lexèmes, dans le lexique, d'autre part, une classification de ces lexèmes tels qu'attestés en corpus.

proche plus encore du lexique espagnol AnCora-Nom (Peris *et al.*, 2010), qui recense les nominalisations présentes dans le corpus Ancora-Es (Taulé *et al.*, 2008) et inclut des informations tant sur la structure argumentale que sur l’aspect lexical des noms. Le lexique Nomage est à notre connaissance le premier à rassembler pour le français des données d’ordre syntaxique (description de la structure argumentale des verbes et des noms), sémantique (caractère massif/comptable des noms, classe aspectuelle de chaque verbe et nom déverbal) et morphologique (mise en relation des bases et des dérivés permettant la comparaison de leurs propriétés linguistiques et distribution par suffixe de dérivation), l’ensemble reposant sur une exploration en corpus. Le volume relativement faible du lexique – 679 verbes et 736 noms décrits – s’explique par la méthodologie employée et le nombre important d’informations associées à chaque lexème.

Dans cet article, nous présenterons tout d’abord les principales spécificités de la ressource Nomage (section 2). Nous nous concentrerons ensuite sur les différentes informations disponibles dans cette ressource et sur la méthode de description de ces informations (section 3). Enfin, les données collectées au cours du projet feront l’objet d’une analyse pour tenter d’apporter un nouvel éclairage aux questions théoriques sous-jacentes, notamment celle de l’héritage des propriétés aspectuelles du verbe source par les noms déverbaux décrits dans la ressource Nomage (section 4).

2. Spécificités de la ressource : un lexique adossé à un corpus annoté

La ressource Nomage a été développée manuellement dans le cadre d’un projet de recherche consacré aux noms déverbaux, notamment à l’étude de leurs propriétés aspectuelles. La méthode générale de développement de la ressource a consisté à extraire les phrases du corpus contenant des formes déverbales et à identifier les unités lexicales correspondant à ces formes.

2.1. Les déverbaux dans le corpus

Le corpus utilisé pour l’annotation des nominalisations est le French TreeBank, corpus du français constitué d’articles du journal *Le Monde* (environ 500 000 mots dans la partie catégorisée) et proposant différents niveaux d’annotation linguistique : marquage morphosyntaxique des lemmes (adjectif, nom commun, préposition, etc.), identification des constituants (groupe nominal, groupe prépositionnel, etc.) et indication de leur fonction dans la phrase (sujet, objet, etc.) (Abeillé *et al.*, 2003).

Sur les 11 800 nominalisations candidates extraites du corpus à partir de critères formels, seules 4 042 ont été analysées dans le cadre du projet Nomage. Nous avons tout d’abord ignoré les nominalisations ne correspondant pas à la tête du syntagme dans lequel elles figuraient (*un permis de construction*), position qui rendait en effet difficilement applicables les tests (cf. *infra*, § 3.1.3). Par ailleurs, certaines nominalisations extraites correspondaient à des désadjectivaux (le lexème INDULGENCE, dérivé

de INDULGENT) et se situaient donc en-dehors du cadre de notre étude. Ensuite, parmi les candidats extraits figuraient beaucoup de conversions (VOYAGE-VOYAGER) également laissées de côté en raison des questions relatives à l'orientation de la dérivation. Enfin, les noms ayant été extraits automatiquement en fonction de la présence d'une chaîne de caractères pouvant être un suffixe (par exemple *-tion*, *-age*, etc.), et malgré l'exigence d'une longueur minimale de caractères précédant ce suffixe, un certain nombre d'entre eux se sont avérés ne pas être des nominalisations (SARCOPHAGE). Une liste d'exclusion (*stop-list*) a été dressée pour exclure automatiquement la plupart des cas susmentionnés, les annotateurs se chargeant d'identifier les erreurs restantes.

2.2. Les déverbaux et leur base verbale dans le lexique

Une fois l'ensemble des formes déverbaux identifié dans le corpus, il nous a fallu dresser la liste des unités à décrire dans le lexique, le plus délicat pour cette tâche étant d'identifier les cas de polysémie.

2.2.1. Nomenclature du lexique

Les entrées du lexique correspondent à des **lexèmes** (ou unités lexicales), entendus ici comme des unités abstraites associées à un ensemble de mots-formes (Lyons, 1977) – par exemple, CONSTRUCTION, associé aux mots-formes *construction* et *constructions* – et étant, d'un point de vue sémantique, monosémique (Mel'čuk *et al.*, 1995 ; Fradin et Kerleroux, 2003). Les unités décrites ne sont donc pas directement les mots (angl. *word*) mais bien les mots pris dans chacun de leur sens (angl. *word sense*). Nous avons de fait opté pour un traitement discret de la polysémie, consistant à décrire dans des entrées distinctes chacune des différentes acceptions de noms polysémiques rencontrés dans notre corpus. Par exemple, aux deux sens de INSTALLATION correspondent les deux lexèmes INSTALLATION#1 (sens actionnel : *L'installation de ce système d'alarme prendra environ deux heures*) et INSTALLATION#2 (sens objectal : *Cette installation tiendra au mieux deux ans*). Suivant la terminologie de la Lexicographie Explicative et Combinatoire (Mel'čuk *et al.*, 1995), nous parlerons de **vocable** pour désigner l'élément regroupant des lexèmes ayant le même signifiant et au moins une composante sémantique commune (comme INSTALLATION#1 et INSTALLATION#2)

Les lexèmes nominaux du lexique sont des nominalisations, c'est-à-dire des noms dérivés par suffixation⁴ de verbes, et qui partagent les propriétés distributionnelles et sémantiques par défaut de la catégorie nominale (Fradin, 2011). L'ensemble des cas de conversion (MARCHE-MARCHER) se trouvent ainsi exclus de notre champ d'étude, pour des raisons à la fois théoriques et méthodologiques. D'une part, la question de l'orientation de la dérivation en cas de conversion (verbe dénominal *vs* nom déverbal) est difficile à trancher systématiquement (Tribout, 2010). D'autre part, le repérage automatique des noms supposément convertis à partir de bases verbales est souvent

4. Les suffixes concernés sont les suivants : *-ade*, *-age*, *-ance*, *-ée*, *-ence*, *-ion*, *-ment*, *-ure*.

impossible, faute précisément de marque formelle de type affixal. Du point de vue sémantique, la plupart des nominalisations décrites dans le lexique Nomage sont des nominalisations au sens restreint, c'est-à-dire des noms construits qui partagent les propriétés de référence abstraite de leurs bases verbales (CÉLÉBRER-CÉLÉBRATION). On trouve toutefois également des nominalisations au sens large, c'est-à-dire des noms qui, quoique dérivés d'un verbe, ne permettent pas de référer directement à ce que dénote ce verbe (GARER-GARAGE).

Le lexique propose également une description des lexèmes verbaux correspondant à la base morphologique des nominalisations décrites. L'identification du verbe source peut poser trois types de difficultés. Tout d'abord, il n'est pas aisé d'établir précisément l'origine d'une nominalisation lorsque celle-ci dérive d'un verbe ayant une forme causative et une forme réfléchie. Par défaut, les nominalisations de ce type apparaissent dans le lexique Nomage avec les deux formes sources possibles (ABONNER/S'ABONNER → ABONNEMENT). Par ailleurs, il est difficile de déterminer, pour certains lexèmes nominaux qui combinent plusieurs étapes de construction, s'ils dérivent d'un verbe déjà construit (REFORMULER → REFORMULATION) ou d'un nom déverbal (FORMULER → FORMULATION → REFORMULATION). Nous optons pour la première solution, considérant que dans les cas en question, la base est le verbe construit, si celui-ci est attesté. Enfin, l'identification de l'origine des nominalisations dénotant un objet pose problème, car il n'est pas facile de déterminer si la nominalisation provient du verbe ou du sens actionnel de la nominalisation. Par exemple, INSTALLATION#2 (objet) provient-il de INSTALLATION#1 (action) ou directement de INSTALLER (Melloni et Bisetto, 2010)? Nous considérons que dans ce genre de cas la base est verbale, la nominalisation avec le même sens que le verbe source n'étant pas toujours attestée dans notre lexique.

2.2.2. Point sur la polysémie

La polysémie des noms déverbaux – dont le faible degré dans le lexique Nomage (1,14 lexèmes par vocable) s'explique par le fait que le corpus utilisé est assez restreint (500 000 mots) et relativement spécialisé (corpus journalistique) – a deux sources principales. Elle peut être tout d'abord héritée du verbe source. Par exemple les deux lexèmes PROMOTION (*sa promotion au poste de directeur financier/la promotion de son livre*) sont dérivés de deux verbes PROMOUVOIR distincts. L'autre source de polysémie est propre au nom : on relève dans le lexique un certain nombre de relations de métonymie entre une action et un des participants de l'action (souvent son résultat) (Melloni et Bisetto, 2010). Le lexique compte par exemple deux lexèmes INSTALLATION, l'un dénotant le fait d'installer quelque chose (INSTALLATION#1), l'autre la chose installée (INSTALLATION#2). Insistons toutefois sur le fait que seules les polysémies avérées dans le corpus sont décrites dans le lexique. Par exemple ADMINISTRATION a deux sens (un sens actionnel et un sens objectal), mais dans le corpus seule l'acception OBJET est présente. Par conséquent, seule l'unité lexicale correspondant au sens OBJET sera décrite dans le lexique. Ce choix, qui peut sembler contestable, a pour but de garantir qu'aucun lexème ne soit décrit uniquement à partir d'exemples construits ou de nos intuitions. Le lexique Nomage n'est qu'une pre-

mière tentative d'annotation sémantico-aspectuelle en corpus, qui présente de ce fait un biais inhérent au choix du corpus. Le biais sous-jacent à ce choix méthodologique pourra être corrigé par la suite en étendant et en diversifiant le corpus utilisé, et par conséquent la couverture du lexique.

Soulignons ici que certains lexèmes sont sémantiquement complexes. Ce phénomène, dont les différents aspects ont déjà été décrits dans la littérature (Pustejovsky, 1995 ; Godard et Jayez, 1996 ; Kleiber, 1999 ; Milicévic et Polguère, 2010), demeure souvent difficile à diagnostiquer lors de la description effective des données. Lorsqu'un signifiant se voit associer deux signifiés, la question se pose de savoir si on est en présence de deux lexèmes (polysémie) ou d'un lexème unique présentant deux facettes sémantiques (Cruse, 2003) et correspondant à un type sémantique complexe, ou type pointé (Pustejovsky, 1995). Dans le second cas, les lexèmes peuvent faire l'objet d'une coprédication, comme dans *Son exposé fut long et ennuyeux*, où *long*, qualifiant le déroulement de l'exposé, s'applique à l'aspect accomplissement d'*exposé*, tandis que *ennuyeux*, qualifiant le contenu informationnel, s'applique à l'acception objet. Un tel cas reçoit dans notre lexique l'étiquette ACC•OBJ.

Le tableau 1 ci-dessous donne un aperçu de la constitution du lexique.

Lexèmes nominaux	746
Vocables nominaux	656
Polysémie nominale	1,14
Lexèmes verbaux	679
Vocables verbaux	648
Polysémie verbale	1,04

Tableau 1. *Constitution du lexique Nomage*

Comme indiqué plus haut, le lexique Nomage ne représente à ce jour qu'un volume relativement faible de données, qui limite dans l'immédiat toute exploitation quantitative des données sur les propriétés aspectuelles des noms décrits. Toutefois, ces descriptions sont d'une densité importante et représentent, d'après nos estimations, environ 1 homme/heure par occurrence décrite.

2.3. Description d'un déverbal

L'information disponible dans la ressource Nomage est illustrée dans les tableaux 2, 3 et 4, avec la description d'AMÉNAGEMENT#1, de son verbe source AMÉNAGER#1 et de ses occurrences dans le corpus. Le nom est décrit ici comme ayant deux arguments (X, celui qui aménage et Y, ce qui est aménagé⁵) et dénote un accomplissement, autrement dit une situation dynamique, durative et culminante (tableau 2).

5. Les arguments n'ont pas reçu de description sémantique dans la version de la ressource présentée dans cet article mais seront caractérisés sémantiquement dans une prochaine version.

id	45
Lexème	AMÉNAGEMENT#1
Structure argumentale	~ de Y par X
Classe aspectuelle	ACC
Occurrences dans le FTB	{id : 1 794 ; id : 1 929}
Base verbale	id : 44

Tableau 2. Description du nom AMÉNAGEMENT#1

Chaque lexème nominal pointe vers le lexème verbal dont il est dérivé (tableau 3). L'entrée verbale contient également une description de la structure argumentale et des propriétés aspectuelles. Les arguments X et Y correspondent respectivement syntaxiquement au sujet et à l'objet direct du verbe, comme l'indique la forme propositionnelle X AMÉNAGER#1 Y.

id	44
Lexème	AMÉNAGER#1
Structure argumentale	X ~ Y
Classe aspectuelle	ACC

Tableau 3. Description du verbe AMÉNAGER#1

Enfin, les lexèmes nominaux pointent vers leurs occurrences en corpus. La description d'une occurrence consiste à décrire la façon dont les arguments du lexème sont réalisés dans la phrase (ici, X n'est pas réalisé tandis que Y l'est sous la forme d'un adjectif relationnel) et à donner le résultat des tests appliqués en corpus, à partir desquels peut être notamment inférée une classe aspectuelle. On notera ici que seule la réalisation syntaxique des arguments du nom est décrite dans la ressource. La réalisation syntaxique des verbes ne l'est pas, les verbes ne faisant pas ici spécifiquement l'objet d'une étude en corpus.

id	1 794
Déverbal	id : 45
Phrase	Tout ce travail préparatoire sera fondamental pour l' aménagement universitaire au cours des cinq prochaines années.
Réalisation des arguments	X : ∅, Y : adj. rel.
Classe aspectuelle inférée des tests	EVT

Tableau 4. Une occurrence de AMÉNAGEMENT#1

3. Description des lexèmes déverbaux et de leurs occurrences en corpus

La ressource Nomage, en plus de fournir une information morphologique sur les nominalisations (suffixes et base verbale), propose une description de leurs propriétés sé-

mantiques (aspect lexical, distinction massif/comptable) et de leur structure argumentale. Ces différentes informations ainsi que la méthode adoptée pour les attribuer sont décrites dans cette section.

3.1. Classification aspectuelle : confrontation de deux méthodes

La principale originalité de la ressource Nomage est d'attribuer une classe aspectuelle à chaque lexème à l'issue d'une double méthode d'annotation : côté lexique, la classification est fondée sur des tests aspectuels appliqués à des exemples construits et menée par des annotateurs formés en sémantique ; côté corpus, la classification est fondée sur une application de tests aspectuels à des phrases du corpus et menée par des annotateurs non linguistes. Ces deux méthodes ont permis, d'une part, de classer les lexèmes selon une classification aspectuelle fine, et, d'autre part, de classer les occurrences de ces mêmes lexèmes de manière empirique (la classification aspectuelle obtenue est alors beaucoup plus rudimentaire et comporte trois grandes classes : EVT, ETAT, OBJET). L'analyse conceptuelle est ainsi validée chaque fois que l'annotation côté corpus concorde avec l'annotation côté lexique. Nous considérons qu'il y a concordance entre les deux annotations dès lors qu'elles n'entrent pas en contradiction l'une avec l'autre, selon la table de correspondance du tableau 5⁶.

Classes aspectuelles côté corpus	Classes aspectuelles côté lexique
EVT	ACC, ACT, ACT-ACC, ACH, (ACC/ACH)-ETAT, (ACC/ACH)●OBJET
ETAT	ETAT, ACC-ETAT, ACH-ETAT
OBJET	OBJET, ACH●OBJET, ACC●OBJET

Tableau 5. Correspondances entre classes aspectuelles du lexique et classes aspectuelles du corpus

Comme le tableau ci-dessus l'indique, seules trois classes aspectuelles sont considérées côté corpus, alors que nous en distinguons onze côté lexique. Si l'on se réfère aux ontologies présentées plus bas figures 1 et 2, les distinctions faites côté corpus reviennent à ne considérer que les sous-nœuds ACTIONS et ÉTATS sous SITUATIONS, qui s'opposent aux OBJETS (non représentés dans les ontologies). Autrement dit, nous cherchons côté corpus à distinguer les OBJETS des SITUATIONS, au sein desquelles nous nous contentons d'une opposition entre SITUATIONS dynamiques (EVT dans le tableau 5, colonne de gauche) vs non dynamiques (ETAT).

3.1.1. Les classes aspectuelles utilisées

La ressource Nomage est le résultat d'une étude qui s'est concentrée essentiellement sur les propriétés aspectuelles des noms déverbaux. Dans leur très grande majorité,

6. Concernant la comparaison chiffrée entre les deux classifications aspectuelles, cf. section 4.3.

les travaux sur l'aspect lexical (ou *Aktionsart*) portent sur le domaine verbal, l'aspect nominal en tant que tel restant peu étudié. Nous présentons ici les classes aspectuelles qui vont nous servir à décrire les nominalisations et leur verbe source côté lexique, en distinguant d'une part les classes aspectuelles communes aux deux catégories grammaticales et d'autre part les classes propres au domaine nominal.

3.1.1.1. Classes aspectuelles communes aux verbes et aux noms

Les quatre premières étiquettes retenues correspondent aux classes vendleriennes, dont nous rappelons ici les principales caractéristiques (Vendler, 1967). Les lexèmes de la classe des états (ETAT) dénotent des situations non dynamiques (CROIRE et ADMIRATION). Parmi les situations dynamiques, on distingue les lexèmes de la classe des activités (ACT), tels que MANIFESTER et PROMENADE, qui dénotent des situations duratives et non culminantes, des accomplissements (ACC) impliquant une situation culminante (RÉPARER et DÉMÉNAGEMENT)⁷ et des achèvements (ACH) dénotant eux aussi une situation culminante mais contrairement aux accomplissements, non durative (DÉCOUVRIR et ACQUISITION). Ces quatre classes aspectuelles sont présentées de manière hiérarchisée dans la figure 1.

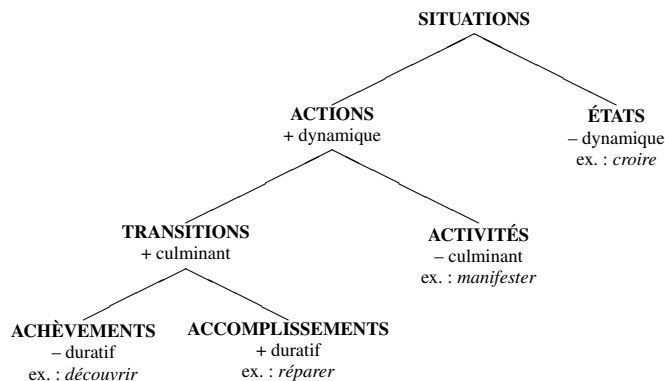


Figure 1. Hiérarchie des classes aspectuelles simples pour les verbes

Nous avons dû recourir à un ensemble de classes aspectuelles moins usuelles pour rendre compte de distinctions sémantiques récurrentes repérées lors de l'annotation des lexèmes étudiés. En effet, certains lexèmes n'entrent dans aucune des classes simples mentionnées ci-dessus, mais forment des catégories intermédiaires : à la charnière entre les achèvements et les états, nous postulons l'existence des « achèvements statifs » (ACH-ETAT), qui répondent favorablement à certains tests réservés aux achèvements mais aussi à certains tests acceptables par les états – notamment des tests de durée, lorsque ceux-ci portent sur un état résultant. C'est le cas, par exemple, de EMPRISONNEMENT qui dénote un achèvement (la mise en prison) suivi d'un état (la période de détention). De la même manière, il existe des « accomplissements statifs » (ACC-ETAT), qui répondent favorablement à certains tests réservés aux accomplissements mais aussi à certains tests acceptables par les états – notamment des tests de durée, lorsque ceux-ci portent sur un état résultant.

7. Une situation est culminante si sa structure inclut un changement d'état.

ifs » (ACC-ETAT) qui décrivent un accomplissement suivi d'un état. C'est le cas de *INVASION* qui renvoie à l'action durative d'envahissement du pays puis à l'état d'occupation du territoire envahi. Constituant une classe intermédiaire entre les ACT et les ACC, il existe des « activités-accomplissements » (ACCgrad), qui dénotent des activités dont chaque étape pourrait être considérée comme le stade final. Tel est le cas de *REFROIDIR*, *RÉTRÉCISSEMENT*, etc. Cette catégorie est également connue sous le nom de *degree achievements* (Dowty, 1979).

3.1.1.2. Classes propres aux noms

Les classes que nous venons de présenter s'appliquent à la fois aux lexèmes verbaux et nominaux. Cependant, l'existence de particularités sémantiques dans le domaine nominal rend indispensable l'ajout de plusieurs catégories aspectuelles, permettant d'étiqueter plus finement les nominalisations. Les verbes d'activité (*JARDINER*, *SE PROMENER*, *MANIFESTER*) ne dérivent pas une classe homogène de nominalisations (Flaux et Van de Velde, 2000 ; Haas *et al.*, 2008 ; Heyd et Knittel, 2009). En effet, l'opposition massif/comptable distingue, au niveau aspectuel, deux types de noms dérivés de verbes d'activité : les nominalisations comptables (*PROMENADE*) et les nominalisations massives (*JARDINAGE*). Du point de vue aspectuel, ces noms décrivent tous des situations dynamiques, duratives et non culminantes, mais seuls les noms comptables dénotent des actions temporellement délimitées, *i.e.* des événements (Haas et Huyghe, 2010). L'étiquette ACT est conservée pour ces déverbaux d'activité comptables, qui sont statistiquement les plus représentatifs de la catégorie, tandis que les déverbaux d'activité massifs se voient assigner l'étiquette HAB (pour « habitude »), en vertu de leur capacité à dénoter des activités routinières (Barque *et al.*, 2009).

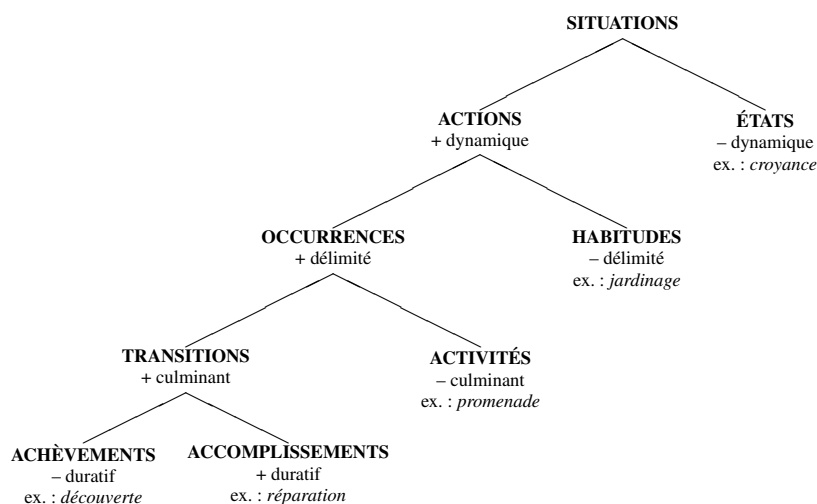


Figure 2. Hiérarchie des classes aspectuelles simples pour les noms

Une autre particularité des noms est que, contrairement aux verbes, ils peuvent dénoter des objets, auquel cas ils sont dépourvus de traits aspectuels. Cette propriété

est connue pour les nominalisations exprimant le résultat d'une action (Grimshaw, 1990), mais elle peut être étendue à un ensemble plus large de nominalisations. Nous envisageons ainsi l'existence d'une classe OBJET, à l'intérieur de laquelle on peut distinguer des noms dénotant des objets matériels (CONSTRUCTION), des noms dénotant des objets dotés d'un contenu informationnel (AFFIRMATION) ou encore des noms dénotant des entités déclenchant un état psychologique (OBSESSION). Enfin, nous avons recours à des classes complexes qui incluent les lexèmes nominaux susceptibles de dénoter une situation, un objet, ou les deux à la fois (Pustejovsky, 1995 ; Godard et Jayez, 1996 ; Milicévic et Polguère, 2010) (cf. section 2.2.2).

Le tableau 6 ci-dessous liste l'ensemble des classes aspectuelles (CA) utilisées pour caractériser sémantiquement les lexèmes (verbaux et nominaux) figurant dans le lexique Nomage. Cette classification étant très fine, elle sera simplifiée pour annoter sémantiquement les occurrences des nominalisations en corpus, comme nous le verrons dans la section suivante.

CA verbales	ACC, ACC-ETAT, ACH, ACH-ETAT, ACT, ACT-ACC, ETAT
CA nominales	CA verbales + ACC●OBJET, ACH●OBJET, HAB, OBJET

Tableau 6. *Classes aspectuelles côté lexique : classification fine*

3.1.2. Classification à partir de phrases contruites

Côté lexique, la classe aspectuelle des lexèmes verbaux et nominaux est attribuée de façon traditionnelle, à partir d'exemples construits⁸.

Les premiers tests appliqués pour attribuer une classe aspectuelle aux lexèmes verbaux reposent sur l'opposition entre verbes d'état et verbes dynamiques, révélés par des tests de dynamicité comme *être en train de V*. Parmi les verbes dynamiques, nous distinguons les verbes culminants (accomplissements et achèvements) des verbes non culminants (activités) grâce au test *en x temps* et au test du paradoxe imperfectif. Enfin, le test *finir/terminer de V* nous permet de distinguer, parmi les verbes culminants, ceux qui sont duratifs (les accomplissements), de ceux qui ne le sont pas (les achèvements).

Dans le domaine nominal, les premiers tests appliqués distinguent les noms qui dénotent une occurrence (au sens aspectuel, cf. *supra* figure 2), acceptant d'entrer dans la construction *N a eu lieu* de ceux qui ne dénotent pas une occurrence (les états

8. La documentation détaillée des tests appliqués en corpus est disponible à l'adresse suivante : <http://nomage.recherche.univ-lille3.fr/> (délivrables→pièces jointes→documentation_projet_nomage.pdf). Les tests permettant de mettre en évidence les propriétés aspectuelles des verbes sont largement répandus dans la littérature (Garey, 1957 ; Vendler, 1967 ; Comrie, 1976 ; Dowty, 1979 ; Lamiroy, 1987 ; Gross, 1996) entre autres. Une part importante du travail théorique préalable à la réalisation de la ressource Nomage a consisté en l'adaptation de ces tests aux prédicats nominaux (Huyghe et Marín, 2008 ; Haas *et al.*, 2008 ; Barque *et al.*, 2009 ; Haas, 2009 ; Barque *et al.*, 2012), *inter alia*.

et les habitudes). Si le nom testé n'est pas une occurrence, il faut déterminer s'il correspond à un état ou à une habitude, ces derniers étant compatibles avec le test *faire du N*. Si au contraire on est en présence d'un nom occurrence, il faut déterminer si cette occurrence est durative (accomplissement ou activité) ou non (achèvement) en appliquant de manière croisée divers tests de durativité (*N en cours*, *N dure x temps*, *commencer/continuer le N*, *N se déroule en/pendant x temps*). Enfin, parmi les noms occurrence duratifs, les tests *le N a été mené à bien*, *en x temps* et le test du paradoxe imperfectif adapté aux noms permettent de distinguer les accomplissements, culminants, des activités, non culminantes (Haas *et al.*, 2008).

3.1.3. Classification à partir de phrases attestées

Une des particularités du projet Nomage est, rappelons-le, d'étudier les nominalisations dans leur contexte naturel d'apparition et non uniquement à partir d'exemples construits. Les tests choisis pour mettre en évidence certaines de leurs propriétés sémantiques (en l'occurrence leurs propriétés aspectuelles et leur caractère massif ou comptable) ont été appliqués par des annotateurs non linguistes auxquels il était demandé d'évaluer le caractère bien formé des phrases transformées selon les tests⁹. Afin de ne pas influencer les annotateurs, ces tests leur étaient présentés de manière non structurée, comme illustré dans la figure 3, qui reprend la fiche (annotée) d'une occurrence du lexème ÉVALUATION.

<p><i>L'évaluation faite selon les critères du BIT (Bureau international du travail) n'est pas plus rassurante.</i></p>
<p>T1_Plusieurs : yes → <i>Plusieurs évaluations, faites selon les critères, ...</i> T2_Avoir lieu : yes → <i>L'évaluation qui a eu lieu hier et qui a été faite selon ...</i> T3_Éprouver/ressentir : no T4_Un peu de : no T5_Durer x temps : yes → <i>L'évaluation, qui a duré 2 jours et qui a été faite ...</i> T6_Se trouver (qq part) : yes → <i>L'évaluation faite selon les [...], et se trouvant actuellement sur le bureau du directeur, n'est pas plus rassurante.</i> T7_Effectuer/procéder : yes → <i>L'évaluation effectuée hier selon les critères ...</i> T8_État de : no T9_Se dérouler : yes → <i>L'évaluation qui s'est déroulée hier ...</i> T10_Card : yes → <i>Deux évaluations, faites selon les critères ...</i></p>

Figure 3. Annotation d'une occurrence du nom ÉVALUATION

9. Pour une description détaillée de l'application des tests en corpus, voir la documentation disponible à l'adresse suivante : http://nomage.recherche.univ-lille3.fr/delivrables-pieces-jointes-guide_annotation_nomage.pdf.

Les trois tests consistant à remplacer l'article de la nominalisation par *plusieurs* (test 1), *un peu de* (test 4) et par un cardinal (test 10) permettent de décider du caractère comptable ou massif d'une occurrence. Dans notre exemple, seuls les tests 1 et 10 fonctionnent, ce qui permet de conclure que cette occurrence du nom ÉVALUATION présente le trait [+comptable]. Les sept autres tests mettent en évidence les principales propriétés aspectuelles de l'occurrence. Les tests consistant à modifier la nominalisation par une relative avec *avoir lieu* (test 2) et *effectuer/procéder* (test 7) indiquent, lorsqu'ils sont positifs, que la nominalisation testée est un événement, ce qui est le cas d'ÉVALUATION. Le test *se dérouler* (test 9) détermine dans ce cas si l'événement est duratif. Du côté des non-événements, les tests *éprouver/ressentir* (test 3) et *état de* (test 8) permettent de détecter les états. Le nom ÉVALUATION étant un événement duratif, il est attendu qu'il ne passe pas ces deux derniers tests. Enfin, le test *se trouver (qq part)* (test 6) permet d'identifier les objets.

Une fois les tests appliqués aux 4 042 occurrences de déverbaux, les résultats sont interprétés selon trois grilles de référence : celle des événements (EVT), qui implique au moins que le test *avoir lieu* soit positif, celle des états (ETAT), dans laquelle au moins l'un des tests 3 ou 8 doit être applicable et enfin celle des objets (OBJET), qui implique que le test 6 s'applique à la nominalisation traitée. Les trois occurrences de déverbaux suivantes se verront ainsi attribuer respectivement les classes EVT•OBJET, OBJET et ETAT en fonction du résultat obtenu aux dix tests présentés dans le tableau 7.

- (1) a. *L'évaluation faite selon les critères du BIT (Bureau international du travail) n'est pas plus rassurante.* (EVT•OBJET)
- b. *Il s'agit de produits récupérés dans les installations de traitement des gaz.* (OBJET)
- c. *Le mécontentement est de plus en plus grand en Pologne à la suite des fortes hausses des prix du gaz, de l'électricité et de l'eau chaude appliquées au début de l'année.* (ETAT)

	(1a) = évaluation	(1b) = installation	(1c) = mécontentement
2. <i>Avoir lieu</i>	yes	no	no
3. <i>Éprouver</i>	no	no	yes
5. <i>Durer x temps</i>	yes	no	yes
6. <i>Se trouver</i>	yes	yes	no
7. <i>VSup dyn.</i>	yes	no	no
8. <i>État de</i>	no	no	no
9. <i>Se dérouler</i>	yes	no	no
Classe inférée	EVT•OBJET	OBJET	ETAT

Tableau 7. *Interprétation des résultats pour la classe aspectuelle*

Les classes aspectuelles attribuées aux occurrences de nominalisations en fonction du résultat des tests sont donc moins fines que celles qui sont attribuables de façon

classique (c'est-à-dire à partir d'exemples construits) aux nominalisations (cf. *supra*, figure 5). Ce décalage s'explique essentiellement par la difficulté d'appliquer certains tests dans des contextes de phrases réelles.

3.2. Description de la structure argumentale

Outre les propriétés aspectuelles des nominalisations, la ressource Nomage propose une description de leurs arguments sémantiques (section 3.2.1) et de la façon dont ces arguments sont réalisés (ou non) syntaxiquement dans les phrases du corpus (section 3.2.2).

3.2.1. Identification des arguments des lexèmes prédicatifs

Le lexique fournit, pour chaque lexème prédicatif (nominal ou verbal), la liste de ses arguments sémantiques, autrement dit des différents actants requis pour définir la situation dénotée par ce prédicat. Ces arguments sémantiques sont représentés dans le lexique par des variables (X, Y, Z, ...). Le lexème AUTORISER est ainsi décrit par X AUTORISER Y à Z.

Les arguments d'une nominalisation correspondent en général à ceux du verbe dont elle dérive¹⁰, cf. 2a par rapport à 2b : X correspond à *la personne qui expédie*, Y à *ce qui est expédié* et Z à *l'endroit où Y est expédié* (ou à *la personne se trouvant à cet endroit*). Pour les besoins de notre exposé, nous avons ici recours à un exemple construit en 2b, néanmoins, notons que la réalisation en surface des arguments de la nominalisation est souvent optionnelle : certains arguments, en particulier l'agent, restent généralement implicites. Ainsi, dans les occurrences extraites du corpus, seul l'argument Y de EXPÉDITION est réalisé en surface, cf. 2b vs 2c.

- (2) a. X EXPÉDIER Y à Z (Ex. *On l'a expédié au Mozambique en novembre 1992.*)
 b. EXPÉDITION de Y à Z par X (Ex. *L'expédition de livres à l'autre bout du monde par Amazon a permis de banaliser le commerce en ligne.*)
 c. EXPÉDITION de Y à Z par X (Ex. *Rien n'a été fait pour faciliter les expéditions de produits agricoles.*)

Il arrive cependant que les différents participants de la situation dénotée ne soient pas identifiés par la même variable dans l'entrée verbale et dans l'entrée nominale, comme c'est souvent le cas avec les verbes causatifs d'état. Dans l'exemple ci-dessous, *la personne satisfaite* est représentée par Y dans l'entrée verbale et par X dans l'entrée nominale. Inversement, *ce qui satisfait* est représenté par la variable X pour le verbe et par la variable Y pour le nom¹¹.

10. L'identification des arguments des verbes du lexique a été facilitée par la consultation de ressources syntaxiques libres comme Dicovalence (Van den Eynde et Mertens, 2003) et Lefff (Sagot, 2010).

11. Signalons qu'en 3b l'argument Y prend la forme d'une proposition infinitive, ce qui constitue une réalisation un peu particulière pour cette structure argumentale. Au vu du caractère

- (3) a. X SATISFAIRE Y (Ex. *Précédant "Rencontres", on retrouve toujours "Racines", l'émission historique conduite par Gérard Noiriel et qui fait l'unanimité (selon le même sondage IFOP), et "Relais", qui satisfait les téléspectateurs fidèles, bien qu'ils trouvent cette séquence un peu trop brève.*)
- b. SATISFACTION de X de Y (Ex. *Alain Dutoya, de son côté, ne cache pas sa satisfaction d'avoir fait aboutir un projet qui, à l'évidence, lui tenait à cœur.*)

Les variables devant être typées sémantiquement dans une prochaine version de la ressource, nous pourrions indiquer explicitement ces cas de divergence.

3.2.2. Encodage de la réalisation syntaxique des arguments

La ressource Nomage propose également une description de la réalisation syntaxique de surface des arguments des lexèmes décrits, dans les phrases du corpus dans lesquelles ils apparaissent¹². L'encodage de la réalisation syntaxique des arguments s'inspire principalement des descriptions produites dans le cadre de la Lexicologie Explicative et Combinatoire (Mel'čuk *et al.*, 1995). Nous distinguons en outre les groupes nominaux compléments employés avec déterminant (Gdét), comme dans l'exemple (4) ci-dessous, des groupes nominaux compléments employés sans déterminant (GN), comme dans l'exemple (5), cette information nous semblant en effet pertinente pour l'analyse de l'aspect.

- (4) Occurrence du lexème « ANNULATION de Y par X » :
Le groupe public français Bull a confirmé l'annulation par une juridiction fédérale américaine du contrat de 4 milliards de francs que sa filiale Zenith avait emporté avec l'US Air Force. ⇒ **X = par Gdét, Y = de Gdét**
- (5) Occurrence du lexème « CONSOMMATION de Y par X » :
On a vu s'affirmer une consommation chinoise de bijoux, qui devrait s'accroître d'ici à la fin du siècle. ⇒ **X = adj rel, Y = de GN**
- (6) Occurrence du lexème « COMMERCIALISATION de Y par X » :
IBM devient ainsi actionnaire de Dassault systèmes à hauteur de 10 % et assure la commercialisation de ses logiciels Catia. ⇒ **X = ∅, Y = de Gdét, Vsup = X assurer dét N**

Comme le montre l'exemple de COMMERCIALISATION (6), seuls sont encodés les arguments réalisés qui dépendent syntaxiquement de la nominalisation. Lorsqu'un argument est présent dans la phrase mais lié par un verbe support à la nominalisation,

régulier de ces réalisations dans notre corpus et des effectifs relativement restreints d'occurrences annotées, nous avons toutefois choisi de les conserver.

12. La ressource Nomage ne décrit donc pas l'ensemble des réalisations syntaxiques possibles de chaque argument d'une nominalisation donnée, mais seulement la réalisation des arguments dans les phrases du corpus dans lesquelles apparaît la nominalisation en question.

il est encodé au niveau de l'attribut VSup : l'argument agentif (*IBM*) n'est pas un dépendant syntaxique de cette occurrence de COMMERCIALISATION mais un dépendant syntaxique du verbe support *assurer*, duquel dépend également la nominalisation.

4. Analyse des données

Les données disponibles dans la ressource Nomage peuvent faire l'objet de différents types d'analyses linguistiques. Nous nous concentrerons ici sur les trois questions suivantes.

- 1) D'un point de vue morphologique, ces données apportent-elles un éclairage nouveau sur le rapport entre la classe aspectuelle d'une nominalisation et son suffixe ?
- 2) Confirment-elles, sur le plan sémantique, l'hypothèse de l'héritage aspectuel dans le processus de nominalisation ?
- 3) Les deux méthodes utilisées pour mettre en évidence les propriétés sémantiques des nominalisations produisent-elles des résultats concordants ?

4.1. Suffixes et classes aspectuelles

Du point de vue morphologique, on peut se demander s'il existe un lien entre la classe aspectuelle d'un lexème déverbal et son suffixe¹³. Le tableau 8 indique la répartition en classes aspectuelles par suffixes : le lexique Nomage compte par exemple 56 lexèmes formés avec le suffixe *-age* et parmi eux, 45 sont des événements, 2 sont des états, etc.

	EVT	ETAT	OBJ	HAB	Total
<i>-ade</i>	6	-	-	-	6
<i>-age</i>	45	2	7	2	56
<i>-ance/-ence</i>	10	19	8	2	29
<i>-ée</i>	13	2	3	-	18
<i>-ion</i>	336	36	61	12	445
<i>-ment</i>	133	12	14	3	162
<i>-ure</i>	11	1	6	2	20
Total	554	72	99	21	736

Tableau 8. Distribution des classes aspectuelles des lexèmes par suffixe

Ce tableau met en évidence que le suffixe le plus représenté est *-ion* (60,5 %), suivi de *-ment* (20,7 %) et de *-age* (7,6 %), ces proportions correspondant à celles avancées dans (Tanguy et Hathout, 2002). Pour ce qui est des classes aspectuelles, EVT est la

13. Les données chiffrées présentées dans les sections 4.1 et 4.2 reposent sur le nombre de lexèmes (analyse côté lexique) et non sur le nombre d'occurrences de ces lexèmes (analyse côté corpus).

classe la plus fréquente (75,3 %), suivie par OBJ (13,5 %) et ETAT (9,8 %).

Concernant la relation entre suffixe et classe aspectuelle, le suffixe *-ance/-ence* est le seul formant de façon majoritaire des états (Gaeta, 2002) ; toutefois, nos résultats montrent également qu'il n'est pas cantonné à cette seule classe, conformément aux observations de (Dal et Namer, 2010). Nos données corroborent d'ailleurs les conclusions de ces dernières, qui distinguent deux cas de figure selon que le nom est formé directement à partir d'une base verbale (ACCOUTUMER→ACCOUTUMANCE) ou selon qu'il est formé à partir d'un adjectif, formé lui-même sur un verbe (IGNORER→IGNORANT→IGNORANCE). Les auteurs montrent que les noms formés sur le premier patron sont susceptibles de manifester toutes les valeurs aspectuelles possibles, et même de dénoter des objets concrets. Les noms formés sur le second patron dénotent, eux, principalement des états.

De leur côté, les suffixes *-age*, *-ée*, *-ment* et *-ion* semblent former un groupe relativement homogène : entre 70 % et 80 % des noms formés avec ces suffixes sont des événements. Le suffixe *-ure* semble produire essentiellement des noms d'événements (55 %) puis des noms d'objets (30 %). D'après les occurrences étudiées, il s'agit d'ailleurs du suffixe produisant, en proportion, le plus de noms d'objets (30 %). Le suffixe *-ade*, quant à lui, est très peu productif, et seuls des noms d'événements ont pu être décrits.

4.2. Héritage des propriétés aspectuelles entre verbes et noms

Les verbes et les noms déverbaux du lexique Nomage ayant été annotés au niveau aspectuel, il devient ainsi possible d'observer l'héritage aspectuel lors du passage du verbe au nom. Sur 736 paires verbe-nominalisation (V-N), 98 doivent être écartées car le nom est catégorisé comme OBJET, classe sémantique entièrement dépourvue de propriétés aspectuelles. Sur les 639 paires restantes, on observe un héritage aspectuel fidèle (*i.e.* le verbe et le nom appartiennent à la même classe aspectuelle) dans 492 cas, soit environ 77 %. Les 147 autres paires V-N (23 %) présentent des divergences d'aspect qui peuvent être totales (le verbe et le nom n'ont pas la même classe aspectuelle) ou partielles (le nom a en partie seulement le même aspect que son verbe source).

Héritage aspectuel fidèle	492 (77 %)
Divergence aspectuelle totale	80 (12,5 %)
Divergence aspectuelle partielle	67 (10,5 %)
Nominalisation OBJET	98 (non comptabilisés)

Tableau 9. Héritage aspectuel : des verbes aux noms

4.2.1. Divergences aspectuelles totales

Parmi les cas de divergences aspectuelles, 80 paires V-N présentent une divergence aspectuelle totale, dont 16 sont des nominalisations d'aspect HAB. Rappelons que la classe aspectuelle HAB n'existe pas dans le domaine verbal puisqu'elle permet

de distinguer parmi les nominalisations dérivées de verbes d'activité celles qui sont comptables (ACT) de celles qui sont massives (HAB). Il existe donc nécessairement une discordance aspectuelle entre le verbe et la nominalisation (EXPORTER (ACT) → EXPORTATION (HAB)). Les autres cas de divergences totales ne s'expliquent pas par l'existence d'une classe réservée au domaine nominal. Le décalage aspectuel peut être faible (INTERVENIR (ACC) → INTERVENTION (ACT), où il s'agit dans les deux cas d'événements duratifs), ou au contraire important (SE COMPORTER (ACT) → COMPORTEMENT (ETAT), qui dénotent, d'une part, une situation dynamique, et de l'autre une situation statique).

4.2.2. Divergences aspectuelles partielles

On trouve 67 paires V-N présentant un aspect partiellement disjoint. Dans les classes aspectuelles que nous distinguons, certaines sont complexes : d'une part, nous avons les classes ACH-ETAT ou ACC-ETAT qui désignent respectivement des lexèmes dénotant des ACH ou ACC suivis d'un état résultant duratif. D'autre part, nous avons des classes du type ACH•OBJET, ACC•OBJET, etc. qui désignent des types sémantiques complexes, auquel cas le lexème dénote soit un type de situation, soit l'autre, voire les deux à la fois. Parmi les cas de divergences partielles, on distingue deux alternatives. D'une part, le verbe peut appartenir à une classe aspectuelle complexe et la nominalisation à une classe simple correspondant à une partie de la classe complexe du verbe (ACCUSER ACH-ETAT → ACCUSATION ACH). Ce cas de figure concerne 25 paires V-N dans le lexique. D'autre part, il arrive que le verbe appartienne à une classe aspectuelle simple ou complexe et la nominalisation à une classe aspectuelle complexe dont l'une des deux facettes correspond à celle du verbe (TASSER ACC → TASSEMENT ACC-ETAT). Cela concerne 42 paires V-N.

Les 98 nominalisations étiquetées OBJET n'ont aucune caractéristique aspectuelle¹⁴. Dès lors, nous ne pouvons plus parler de divergence d'aspect, mais plutôt d'un changement ontologique : le verbe dénote une situation pourvue de propriétés aspectuelles, la nominalisation dénote une entité concrète qui en est par nature dépourvue.

Classe aspectuelle du verbe	Classe sémantique du nom
EVT culminants (ACH ou ACC) 68	OBJET
EVT non culminants (ACT) 21	
ETAT 9	

Tableau 10. *Origine aspectuelle des OBJET*

Les verbes donnant des nominalisations OBJET dérivent majoritairement de situations pourvues d'un point culminant (ACC, ACC-ETAT, ACC-ACT, ACH et ACH-

14. Les nominalisations du type « classe aspectuelle•OBJET » ont été comptabilisées parmi les cas de divergences aspectuelles partielles puisqu'il y a conservation de l'aspect du verbe et ajout d'une caractérisation sémantique, ex. COTISER (ACH) → COTISATION (ACH•OBJET), AFFIRMER (ACH) → AFFIRMATION (ACH•OBJET).

ETAT confondus). Or il est connu que les prédicats verbaux culminants ont régulièrement une lecture résultative ou concrète (Grimshaw, 1990 ; Pustejovsky, 1995 ; Alexiadou, 2001 ; Osswald, 2005 ; Van de Velde, 2006).

4.3. Comparaison des deux méthodes d'attribution de classe aspectuelle

Deux méthodes d'annotation sémantique ont été appliquées dans ce projet. La première consiste à attribuer une classe aspectuelle à des lexèmes à partir d'exemples construits. La seconde consiste à soumettre les occurrences d'une nominalisation à une série de tests, appliqués par des annotateurs non linguistes. Arrive-t-on aux mêmes résultats en suivant ces deux méthodes ?

Avant d'étudier les données pour mesurer le degré de convergence entre les classes attribuées aux lexèmes et celles attribuées à leurs occurrences, il convient de noter deux points importants. Tout d'abord, les résultats des tests appliqués à chaque occurrence de nominalisation n'ont pas toujours permis de leur attribuer une classe aspectuelle¹⁵. Rappelons ensuite que les distinctions de classes aspectuelles côté lexique (ACC, ACH, etc.) sont plus fines que celles attribuées de façon empirique côté corpus (EVT, ETAT, OBJET). Le tableau 5 présenté plus haut indique les correspondances entre ces deux jeux de classes aspectuelles.

Comme le montre le tableau 11, le degré de non-contradiction entre les étiquettes assignées aux occurrences du corpus et les étiquettes assignées aux lexèmes dans le lexique est relativement élevé. Sur les 2 309 occurrences de nominalisations étiquetées EVT dans le corpus, 2 001 sont des occurrences de lexèmes qui sont bien identifiés comme étant des EVT dans le lexique (autrement dit, des activités, des accomplissements ou des achèvements). Concernant les états, ce sont 136 occurrences sur un total de 217. Enfin, concernant les occurrences d'objets, plus de 90 % d'entre elles sont des occurrences de lexèmes typés OBJET dans le lexique.

Classe aspectuelle	Nombre d'occurrences	Distribution lexique
EVT	2 309	EVT (2 001), ETAT (94), OBJET (153), AUTRE (61)
ETAT	217	ETAT (136), EVT (53), OBJET (22), AUTRE (6)
OBJET	232	OBJET (211), EVT (19), ETAT (0), AUTRE (2)

Tableau 11. Degré de coïncidence entre données du corpus et données du lexique

Les cas de non-coïncidence ont deux sources principales. D'une part, il demeure, à ce stade de l'élaboration de notre ressource (version 1.1), des erreurs d'encodage de la

15. 2 758 occurrences ont pu recevoir une caractérisation aspectuelle par la méthode empirique (c'est-à-dire grâce aux tests appliqués en corpus) sur les 4 042 occurrences extraites.

classe aspectuelle tant au niveau du lexique qu'au niveau du corpus. En effet, l'annotation humaine comporte toujours une part d'aléa : fluctuation des jugements d'acceptabilité, erreurs de transcription des résultats trouvés, oublis, etc. Ainsi, 13 occurrences du lexème PROCÉDURE (dans le sens de *procédure judiciaire*) sont par exemple étiquetées EVT dans le corpus alors que le lexème est décrit comme OBJET dans le lexique : le lexème dénote pourtant une activité et donc un événement. D'autre part, on observe un certain nombre d'erreurs d'attribution d'une occurrence à un lexème : certaines occurrences d'ADMINISTRATION#2 (au sens d'*ensemble des personnes qui administrent*) ont été déclarées comme des occurrences d'ADMINISTRATION#1 (*fait d'administrer*). La confrontation des données annotées en corpus *vs* décrites dans le lexique facilite ainsi la procédure de révision de la ressource Nomage.

5. Conclusion

Dans cet article, nous avons présenté le travail effectué dans le cadre du projet Nomage. Le résultat des phases d'annotation successives a permis la conception d'un lexique électronique de noms déverbaux du français, annotés syntaxiquement et sémantiquement. La ressource lexicale conçue propose notamment une classification aspectuelle des nominalisations, en relation avec celle de leurs bases verbales. La double approche utilisée – côté lexique, sur des exemples construits *vs* côté corpus, sur des exemples attestés – a permis d'obtenir des données stables et utilisables pour de futurs travaux sur l'aspect nominal. Les deux approches se sont alimentées mutuellement : la réflexion théorique a contribué à l'annotation en corpus en fournissant des tests linguistiques et une ontologie, et en retour, l'approche empirique a permis de remettre en question et d'améliorer l'analyse théorique, s'agissant notamment de la pertinence des classes aspectuelles retenues. Les données rassemblées nous permettent ainsi d'apporter des réponses, même partielles, aux questions présentes en filigrane dans l'ensemble du travail.

Ainsi, l'héritage aspectuel entre verbes et noms apparentés semble, dans la grande majorité des cas, être fidèle : d'après nos observations, verbes et noms présentent globalement les mêmes propriétés aspectuelles. Il existe cependant des cas de divergences, qui peuvent être de deux types : soit la catégorie varie, au sein d'un système aspectuel applicable au domaine verbal, soit la catégorie nominale pertinente n'existe pas dans le domaine verbal. Ce dernier cas de figure révèle que les systèmes aspectuels nominal et verbal ne sont pas superposables, et que la catégorie grammaticale détermine la structure de l'aspect lexical (Huyghe, 2011). De nouvelles questions se posent alors. Ainsi, par exemple, dans la mesure où il n'existe pas de verbes dénotant des OBJETS, à quelles conditions un verbe peut-il donner lieu à une nominalisation de ce type (Melloni et Bisetto, 2010) ? Ou encore : le statut massif/comptable, fondamental dans la détermination de l'aspect nominal, correspond-il à une propriété lexicale dans le domaine verbal ?

Sur la question de la concordance entre l'annotation côté lexique et l'annotation côté corpus, en ne considérant que les cas où la double annotation est effective (*i.e.* en

laissant de côté les 1 284 occurrences qui n'ont pas reçu d'étiquette aspectuelle), on obtient un taux de concordance élevé – environ 85 % toutes classes confondues. Ce résultat tend à valider la méthode de double annotation mise en œuvre dans le cadre du projet. Les divergences constatées sont en grande partie dues à des erreurs humaines. Reste le cas des occurrences auxquelles nous n'avons pas pu attribuer de classe aspectuelle à l'issue de l'application des dix tests en corpus (tous marqués négativement). L'absence d'étiquetage tient soit à l'impossibilité d'appliquer les tests en raison de blocages contextuels, soit à des erreurs d'annotation – on relève principalement des cas de rejets indus, concernant notamment les tests « avoir lieu » et « plusieurs ». Ces problèmes devront conduire à reconsidérer la formulation de certains tests.

En ce qui concerne le lien entre le suffixe et l'aspect nominal, l'analyse en corpus a permis de confirmer certaines hypothèses présentes dans la littérature, notamment que les suffixes *-ance/-ence* sont ceux qui ont la plus forte propension à former des ETAT, que *-age*, *-ée*, *-ment* et *-ion* forment majoritairement des EVT, et enfin que *-ure* forme essentiellement des OBJET. D'autres pistes de travail sont à envisager pour la suite. Ainsi pourra-t-on s'intéresser aux cas de conversion (MARCHE-MARCHER), non traités dans Nomage car non détectables automatiquement. De même, les nominalisations d'adjectifs (FIDÈLE-FIDÉLITÉ) et les noms événementiels non construits (CRIME) pourront être soumis à un protocole d'annotation similaire à celui développé dans le projet Nomage.

Le lexique proposé, bien que de taille modeste, offre une information lexicale complète sur les noms déverbaux (analyse aspectuelle, structure argumentale, données morphologiques) qui pourra être utilisée dans le cadre d'études portant sur l'aspect des noms, sur la question de l'héritage aspectuel ou encore sur la réalisation des arguments des prédicats nominaux. Le lexique, encodé dans un format standardisé (LMF), est par ailleurs voué à s'étendre avec la description d'autres unités prédicatives (noms désadjectivaux, noms prédictifs non dérivés) et la mise en commun d'informations provenant d'autres ressources lexicales libres ayant un lien direct avec Nomage (par exemple Verbaction, pour la nomenclature et le Lefff, pour la description des arguments). Cette ressource participe ainsi d'une dynamique actuelle de la communauté linguistique dans le domaine de la constitution de ressources lexicales interopérables. Le lexique Nomage constitue, en outre, un premier pas vers des ressources sémantiquement denses, exploitables pour des problématiques d'annotation syntaxique et sémantique automatique de grands corpus.

6. Bibliographie

- Abeillé A., Clément L., Toussanel F., « Building a treebank for French », in A. Abeillé (ed.), *Treebanks, Building and Using Parsed Corpora*, Kluwer, Dordrecht, 2003.
- Alexiadou A., *Functional Structure in Nominals : Nominalization and Ergativity*, John Benjamins, Amsterdam / Philadelphia, 2001.
- Barque L., Fábregas A., Marín R., « Les noms d'état psychologique et leurs "objets" : étude d'une alternance sémantique », *Lexique*, vol. 20, p. 21-41, 2012.

- Barque L., Huyghe R., Jugnet A., Marín R., « Two types of deverbal activity nouns in French », *5th International Conference on Generative Approaches to the Lexicon*, Pisa, p. 169-175, 2009.
- Chomsky N., « Remarks on Nominalizations », in A. R. P. Rosenbaum (ed.), *Readings in English Transformational Grammar*, Ginn and Co, Waltham (MA), p. 184-221, 1970.
- Comrie B., *Aspect, an Introduction to the Study of Verbal Aspect and Related Problems*, University Press, Cambridge, 1976.
- Cruse D. A., « Aux frontières de la polysémie, les micro-sens », in S. Rémi-Girard, L. Panier (eds), *La polysémie ou l'empire des sens. Lexique, discours, représentation*, PUL, Lyon, p. 131-140, 2003.
- Dal G., Namer F., « Les noms en *-ancel-ence* du français : quel(s) patron(s) constructionnel(s) ? », *Actes en ligne du 2e Congrès Mondial de Linguistique Française*, Nouvelle Orléans, p. 893-907, 2010.
- Dowty D., *Word Meaning and Montague Grammar*, D. Reidel Publishing Co., Dordrecht, 1979.
- Flaux N., Van de Velde D., *Les noms en français : esquisse de classement*, Ophrys, Paris, 2000.
- Fradin B., « Les nominalisations et la lecture 'moyen' », *Lexique*, vol. 20, p. 129-156, 2011.
- Fradin B., Kerleroux F., « Quelles bases pour les procédés de la morphologie constructionnelle ? », in B. Fradin, G. Dal, N. Hathout, F. Kerleroux, M. Plénat, M. Roché (eds), *Les unités morphologiques, Silexicales, Vol.3*, UMR STL 8163-CNRS/Université Lille 3, Villeneuve d'Ascq, p. 78-84, 2003.
- Gaeta L., *Quando i verbi compaiono come nomi. Un saggio di morfologia naturale*, FrancoAngeli, Milano, 2002.
- Garey H., « Verbal aspect in French », *Language*, vol. 33, n° 2, p. 91-110, 1957.
- Godard D., Jayez J., « Types nominaux et anaphores : le cas des objets et des événements », in W. D. Mulder, L. T.-D. Ryck, C. Veters (eds), *Cahiers Chronos 1*, Rodopi, Amsterdam/Atlanta, GA, p. 41-58, 1996.
- Grimshaw J., *Argument structure*, MIT Press, Cambridge, MA, 1990.
- Gross G., « Prédicat nominaux et compatibilité aspectuelle », *Langages*, vol. 121, p. 54-72, 1996.
- Haas P., Comment l'aspect vient aux noms. Les propriétés aspectuelles des noms à l'épreuve des restrictions de sélection imposées par certaines prépositions, PhD thesis, Université Lille 3, 2009.
- Haas P., Huyghe R., « Les propriétés aspectuelles des noms d'activités », *Cahiers Chronos*, vol. 21, p. 103-118, 2010.
- Haas P., Huyghe R., Marín R., « Du verbe au nom : calques et décalages aspectuels », *Congrès Mondial de Linguistique Française (CMLF 2008)*, Paris, p. 2039-2053, 2008.
- Hathout N., Namer F., Dal G., « An experimental constructional database : the MorTAL project », in P. Boucher (ed.), *Many Morphologies*, Cascadilla, Somerville, Mass, p. 179-209, 2002.
- Heyd S., Knittel M.-L., « Les noms d'activité parmi les noms abstraits : propriétés aspectuelles, distributionnelles et interprétatives », *Linguisticae investigationes*, vol. XXXI-1, p. 124-148, 2009.
- Huyghe R., « (A)telicity and the mass-count distinction : the case of French activity nominalizations », *Recherches Linguistiques de Vincennes*, vol. 40, p. 101-125, 2011.

- Huyghe R., Marín R., « L'héritage aspectuel des noms déverbaux en français et en espagnol », *Faits de Langue*, vol. 30, p. 265-273, 2008.
- Kleiber G., *Problèmes de sémantique. La polysémie en question*, Presses Universitaires du Septentrion, Villeneuve d'Ascq, 1999.
- Lamiroy B., « The complementation of Aspectual Verbs in French », *Language*, vol. 63, n° 2, p. 278-298, 1987.
- Lees R. B., *The Grammar of English Nominalizations*, Mouton, The Hague, 1960.
- Lyons J., *Semantics*, Cambridge University Press, Cambridge, 1977.
- Macleod C., Grishman R., Meyers A., Barret L., Reeves R., « NOMLEX : A Lexicon of Nominalizations », *Proceedings of Euralex'98*, Liège, p. 187-193, 1998.
- Melloni C., Bisetto A., « On the Interpretation of Nominals : Towards a Result-Oriented Verb Classification », in O. Souleimanova (ed.), *Sprache und Kognition : Traditionelle und neue Ansätze. Language and cognition : traditional and new approaches. Proceedings of the 40th Linguistics Colloquium, Moscow 2005.*, Frankfurt am Main, p. 165-178, 2010.
- Mel'čuk I., Clas A., Polguère A., *Introduction à la lexicologie explicative et combinatoire*, Duculot, Louvain La Neuve, 1995.
- Milicévic J., Polguère A., « Ambivalence sémantique des noms de communication langagière en français », *Congrès Mondial de Linguistique Française (CMLF 2010)*, EDP Sciences (www.linguistiquefrancaise.org), Paris, p. 1029-1050, 2010.
- Milner J.-C., *Ordres et raisons de langue*, Le Seuil, Paris, 1982.
- Osswald R., « On Result Nominalization in German », *Proceedings of Sinn und Bedeutung 9*, p. 256-270, 2005.
- Peris A., Taulé M., Rodríguez H., « Semantic Annotation of Deverbal Nominalizations in the Spanish corpus AnCora », *Proceedings of The Ninth International Workshop on Treebanks and Linguistic Theories (TLT9)*, Tartu, p. 187-198, 2010.
- Pustejovsky J., *The Generative Lexicon*, MIT Press, Cambridge, 1995.
- Sagot B., « The Lefff, a freely available and large-coverage lexicon for French », *Proceedings of LREC 2010*, La Valetta, Malta, p. 2744-2751, 2010.
- Tanguy L., Hathout N., « Webaffix : un outil d'acquisition morphologique dérivationnelle à partir du web », *Actes de TALN 2002*, Nancy, p. 245-254, 2002.
- Taulé M., Martí M., Recasens M., « Ancora : Multilevel Annotated Corpora for Catalan and Spanish », *Proceedings of 6th International Conference on Language Resources and Evaluation*, Marrakech, p. 96-101, 2008.
- Tribout D., *Les conversions de nom à verbe et de verbe à nom en français*, PhD thesis, Université Paris 7, Paris, 2010.
- Van de Velde D., *Grammaire des événements*, Presses Universitaires du Septentrion, Villeneuve d'Ascq, 2006.
- Van den Eynde K., Mertens P., « La valence : l'approche pronominale et son application au lexique verbal », *Journal of French Language Studies*, vol. 13, p. 63-104, 2003.
- Vendler Z., *Linguistics in Philosophy*, Cornell University Press, Ithaca, N.Y., 1967.