Le projet Nomage
Description des noms déverbaux du français

Équipe Nomage*
STL, Université de Lille 3

31 décembre 2010

Table des matières

1 Introduction **3**

2 Positionnement théorique **3**
2.1 Identification des lexèmes à traiter 3
2.1.1 Types de nominalisations 4
2.1.2 Source morphologique 5
2.2 Définition des classes aspectuelles nominales 6
2.2.1 Classes aspectuelles verbales s'appliquant également aux noms . . . 6
2.2.2 Classes aspectuelles propres aux noms 7
2.2.3 Hiérarchies (verbale et nominale) des classes aspectuelles simples . . 8

3 Le corpus Nomage **9**
3.1 Présentation du corpus 9
3.2 Annotation aspectuelle des nominalisations en corpus 10
3.2.1 Présentation des 10 tests 10
3.2.2 Attribution d'une classe aspectuelle à partir du résultat des tests . . 11
3.2.3 Exemple d'annotation de noms déverbaux 12
3.3 Réalisation syntaxique des arguments dans le corpus 13

*Les membres du projet Nomage ayant contribué au développement des ressources présentées dans ce document sont : Antonio Balvet, Lucie Barque, Marie-Hélène Condette, Pauline Haas, Richard Huyghe, Anne Jugnet, Rafael Marín et Aurélie Merlot. + annotateurs

4 Le lexique Nomage **15**
4.1 Présentation générale : une entrée du lexique Nomage 15
4.2 Quelques chiffres 16
4.3 Structure argumentale des prédicats décrits 17
4.4 Classe aspectuelle d'un lexème 18
4.4.1 Rappel des étiquettes de classes aspectuelles utilisées 18
4.4.2 Tests pour l'attribution de classes aspectuelles aux verbes 18
4.4.3 Tests pour l'attribution de classes aspectuelles aux noms 20

Annexe 1 : Modélisation de la base de données Nomage **22**

Annexe 2 : Réalisation syntaxique des arguments **27**

Annexe 3 : Modélisation de la base de données Nomage **32**

1 Introduction

Le projet Nomage, mené au laboratoire STL de l'université Lille 3[1], vise la description des propriétés aspectuelles des noms dérivés de verbes, description qui repose sur l'observation et l'annotation de ces noms en corpus[2]. L'objectif théorique sous-jacent à ce projet est de voir dans quelle mesure les noms héritent des propriétés aspectuelles des verbes dont ils sont dérivés.

Ce projet a donné lieu au développement de deux ressources :

1. Un **corpus** composé d'un ensemble de phrases extraites du French Treebank contenant toutes au moins un nom déverbal. Chaque nom déverbal y est annoté sémantiquement et syntaxiquement.
2. Un **lexique** des noms déverbaux décrivant uniquement les lexèmes instanciés dans les phrases du corpus. L'objectif du projet Nomage étant l'évaluation de l'héritage aspectuel entre verbes et noms, le lexique décrit également le verbe source de chaque nom décrit.

La suite du document est structurée de la façon suivante : la section 2 présente nos choix théoriques relatifs à la description des noms déverbaux. La section 3 décrit l'annotation des noms déverbaux en corpus (application de tests aspectuels menant à l'attribution d'une classe aspectuelle et indication des arguments des noms réalisés dans la phrase). Enfin, la section 4 est consacrée au lexique Nomage, qui décrit de façon plus générale et plus complète les noms déverbaux rencontrés dans le corpus ainsi que les unités verbales dont ils sont dérivés.

2 Positionnement théorique

2.1 Identification des lexèmes à traiter

Les unités lexicales traitées dans le cadre de Nomage sont des lexèmes, le **lexème** étant défini ici comme une unité abstraite structurée autour d'**un sens** exprimable par un ensemble de mots-formes que seule distingue la flexion (Polguère, 2003 ; Fradin et Kerleroux, 2003). Les unités décrites dans Nomage ne sont donc pas directement les mots (engl. *word*) mais bien les mots pris dans chacun de leur sens (engl. *word sense*)[3].

[1] http ://nomage.recherche.univ-lille3.fr

[2] Le corpus utilisé est le French Treebank (Abeillé, 2003).

[3] Rappelons que nous ne décrivons dans le lexique Nomage que les lexèmes instanciés dans les phrases du corpus.

Nous optons ainsi pour une approche discrète de la polysémie. Les deux lexèmes PROMOTION suivants ont par exemple été identifiés dans le corpus :

> PROMOTION#1 = Nomination à une fonction supérieure (*C'est arrivé après sa promotion au poste de directeur financier.*)
> PROMOTION#2 = Action de provoquer le développement ou le succès de quelque chose (*Chirac va faire la promotion de son livre en plein marasme judiciaire.*)

2.1.1 Types de nominalisations

Les lexèmes déverbaux décrits dans le lexique Nomage relèvent de deux types de nominalisations : celles dont le sens équivaut à celui du verbe auquel elles sont liées et celles dont le sens n'est pas équivalent à celui de leur verbe d'origine (Fradin, itre) :

1. Les **nominalisations au sens restreint** sont des lexèmes nominaux morphologiquement construits à partir d'un prédicat verbal, (i) qui permettent de référer en discours à ce que dénote ce prédicat et (ii) qui partagent les propriétés distributionnelles et sémantiques par défaut de la catégorie N, notamment la possibilité d'être la tête d'un syntagme nominal. Ces nominalisations peuvent ainsi servir d'anaphore au verbe qui leur correspond, comme illustré ci-dessous (exemple de Fradin, itre) :

   *Il parlait de créer de nouveaux emplois mais ces **créations** n'ont jamais vues le jour.*

2. Les **nominalisations au sens large** ont les mêmes propriétés que les nominalisations au sens restreint si ce n'est qu'elles ne permettent pas de référer en discours à ce que dénote leur prédicat d'origine, autrement dit, ne peuvent être paraphrasées par "action de Ver", "fait de Ver", "fait d'être Vé", etc.

Le lexique nomage décrit pour l'essentiel des nominalisations au sens restreint mais aussi des nominalisations n'ayant pas le même sens que leur verbe source, comme illustré ci-dessous avec les deux lexèmes du vocable INSTALLATION, dont seul le premier correspond à une nominalisation au sens restreint.

> ASSOCIATION#1 = fait de (s')associer
> ASSOCIATION#2 = groupement de personnes

Notons que les nominalisations au sens large peuvent figurer dans le lexique même si la nominalisation au sens restreint qui leur correspond n'y figure pas. C'est le cas par exemple d'AGGLOMÉRATION, qui n'apparaît dans le corpus que dans son sens de *ville* et non dans

son sens de *fait d(e s)'agglomérer*. Le lexique décrit donc uniquement le lexème dénotant une ville.

2.1.2 Source morphologique

Un lexème déverbal, quelle que soit la définition qu'on en donne (sens restreint ou sens large) est dérivé morphologiquement d'un lexème verbal. Or l'identification de cette source verbale ne va pas toujours sans poser de problèmes, dont les trois suivants :

• Certains lexèmes peuvent tout d'abord être dérivés aussi bien d'un verbe causatif non réfléchi que de son emploi réfléchi. La solution que nous avons adoptée consiste à avoir des lexèmes nominaux qui, étant dérivés d'un verbe ayant deux emplois possibles (causatif et réfléchi), ont deux structures argumentales possibles[4]. En voici deux exemples :

- ABONNEMENT de Y à Z par X / ABONNEMENT de X à Y
  → **Lexème source :** X ABONNER Y à Z / X S'ABONNER à Y
- AMÉLIORATION de Y par X / AMÉLIORATION de X
  → **Lexème source :** X AMÉLIORER Y / X S'AMÉLIORER

• Un autre cas problématique à traiter est celui des lexèmes nominaux construits, dont on ignore s'ils résultent de la dérivation d'un verbe construit (*reformulation*, dérivé de *reformuler*) ou si au contraire ils ont été construits après dérivation (*reformulation* dérivé de *formulation*, lui-même dérivé de *formuler*). Nous avons décidé de toujours opter, lorsque le verbe construit existe, pour la première solution, comme illustré ci-dessous. Nous n'avons en effet pas voulu ici introduire de biais dans notre étude de l'héritage aspectuel entre noms et verbes.

- REFORMULATION de Y par X
  → **Lexème source :** X REFORMULER Y
- SURÉVALUATION de Y par X
  → **Lexème source :** X SURÉVALUER Y

• Enfin, l'identification de la source sémantique des nominalisations au sens large (celles qui ne dénotent pas la sitation dénotée par le verbe apparenté) pose également problème. Le nom INSTALLATION#2 [*La vétusté des installations y est pour beaucoup*] vient-il du nom INSTALLATION#1 ou du verbe INSTALLER (Bisetto et Melloni, 2008) ? Nous avons décidé d'indiquer dans tous les cas la source verbale, qu'elle soit directe ou indirecte, pour la

[4]Nous verrons plus loin (section 3.3 et annexe 2) comment ces cas sont traités dans la base Nomage lors de l'encodage de la structure argumentale syntaxique.

raison simple que la nominalisation au sens restreint associée à la nominalisation au sens large n'est pas toujours présente dans le lexique (cf. *supra*, l'exemple d'AGGLOMÉRATION).

2.2 Définition des classes aspectuelles nominales

Dans le projet Nomage, la description des propriétés aspectuelles des noms déverbaux et des verbes qui leur sont liés suit deux méthodes, fondées toutes les deux sur l'introspection[5]. La première méthode, appliquée dans le cadre du développement du lexique (cf. section 4), consiste à attribuer une classe aspectuelle au lexème nominal – et à son verbe source – de manière classique, c'est-à-dire en appliquant un certain nombre de tests à partir d'exemples construits. La seconde méthode, liée au développement du corpus (cf. section 3), consiste à appliquer des tests partiellement différents de ceux appliqués sur des exemples construits, notamment parce qu'ils ont été pensés pour être applicables à des "vraies" phrases, en l'occurrence celles du corpus. Nous reviendrons plus loin sur les résultats de la confrontation de ces deux méthodes. Avant cela, il convient de décrire ici le jeu des classes aspectuelles attribuables à une nominalisation. En effet, si les classes aspectuelles des lexèmes verbaux sont relativement bien connues, il nous a fallu caractériser dans le cadre de Nomage les classes pouvant s'appliquer à des lexèmes nominaux.

2.2.1 Classes aspectuelles verbales s'appliquant également aux noms

Les classes aspectuelles utilisées pour catégoriser les unités du lexique Nomage sont pour l'essentiel des classes aspectuelles bien connues dans la littérature. Ces classes valent aussi bien pour les prédicats verbaux que pour les prédicats nominaux.

1. La classe des **états** (ETAT) inclut les lexèmes dénotant une situation non dynamique. Les lexèmes POSSÉDER et ADMIRATION appartiennent par exemple à cette classe.
2. La classe des **activités** (ACT) inclut les lexèmes dénotant une situation dynamique durative mais non culminante. Les lexèmes MANIFESTER et PROMENADE sont des exemples d'activités.
3. La classe des **accomplissements** (ACC) inclut les lexèmes verbaux dénotant une situation dynamique durative culminante. Les lexèmes RÉPARER et DÉMÉNAGEMENT sont des exemples d'accomplissements.
4. La classe des **achèvements** (ACH) inclut les lexèmes verbaux dénotant une situation dynamique culminante mais non durative. Les lexèmes ACQUISITION#1 et

[5]Nous n'avons pas exploré ici de méthode d'attribution automatique de propriétés aspectuelles. Nous espérons cependant que le corpus des nominalisations annotées dans le cadre du projet pourra servir à l'apprentissage d'un programme annotant automatiquement des nominalisations rencontrées en corpus.

ADOPTER en sont des exemples.

Le travail de description qui a donné lieu au développement du lexique nous a toutefois conduit à définir d'autres classes aspectuelles, classes catégorisant là encore aussi bien des verbes que des noms.

1. La classe des **achèvements-statifs** (ACH•ETAT) inclut les lexèmes dénotant un achèvement suivi d'un état borné (par un autre achèvement non inclus dans le sens du lexème). Par exemple, le lexème EMPRISONNEMENT dénote un achèvement (la mise en prison) suivi d'un état qui dure jusqu'à la sortie de prison.
2. La classe des **accomplissements-statifs** (ACC•ETAT) inclut les lexèmes dénotant un accomplissement suivi d'un état borné (par un autre achèvement non inclus dans le sens du lexème). Le lexème INVASION par exemple dénote un accomplissement (le fait d'envahir un pays, qui peut prendre du temps) suivi d'un état (l'occupation du territoire envahi) qui dure jusqu'à la libération du territoire envahi.
3. La classe des **activité-accomplissement** (ACT-ACC) inclut les lexèmes qui dénotent des activités dont chaque étape peut être considérée comme finale (d'où l'aspect accomplissement). Les lexèmes typiques de cette classe sont ceux que l'on appelle dans la littérature des *degree-achievement*, comme REFROIDIR, RÉTRÉCISSEMENT, etc.

2.2.2 Classes aspectuelles propres aux noms

Les classes que nous venons de présenter succintement s'appliquent aux verbes et aux noms. Le domaine nominal présente toutefois des spécificités par rapport au domaine verbal :

Tout d'abord, l'opposition massif/comptable distingue, au niveau aspectuel, deux types de noms dérivés de verbe d'activité :

1. les noms d'activité, comptables. Ex. MANIFESTATION [*Plusieurs manifestations ont eu lieu dans les grandes villes de France.*].
2. les noms d'**habitude** (HAB), qui eux sont massifs. Ex. JARDINAGE [*On recherche des bénévoles pour faire du jardinage.*].

La seconde spécificité des noms par rapport aux verbes réside dans le fait qu'ils peuvent dénoter des objets, c'est-à-dire des entités qui n'ont pas de propriétés aspectuelles.

1. La classe des **objets** (OBJET) inclut des lexèmes des objets concrets, tels que AGGLOMÉRATION, CONSTRUCTION ou encore INSTALLATION. Certains lexèmes de cette

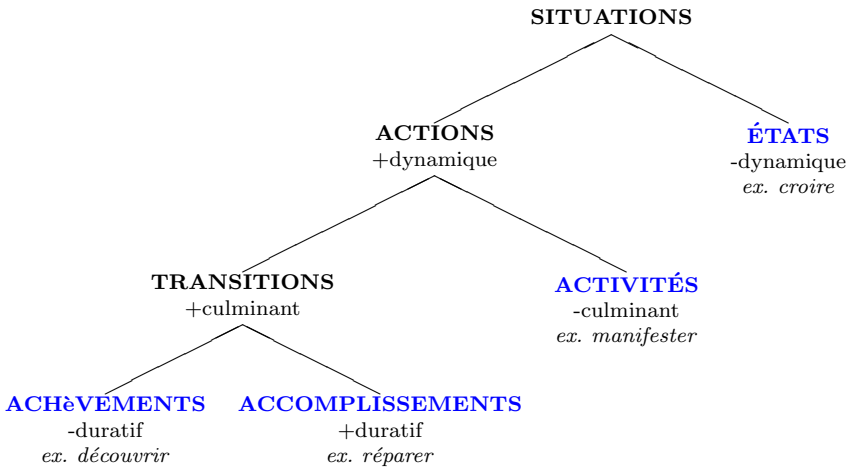


FIG. 1 – Hiérarchie des classes aspectuelles simples pour les verbes

classe ont un type plus spécifique. C'est le cas des "objets informationnels" et des "objets psychologiques" :

(a) La classe des **objets informationnels** (OBJETinfo) inclut des lexèmes dénotant un objet ayant un contenu informationnel (pouvant être lu, par exemple). Ex. ADDITION [*Il a lu l'addition et a fait une drôle de tête.*].

(b) La classe des **objets psychologiques** (OBJETpsy) incluent les lexèmes dénotant quelque chose qui provoque un état psychologique. Par exemple, le nom OBSESSION peut dénoter quelque chose qui est l'objet d'une obsession.

Les **classes complexes** sont des classes qui incluent les lexèmes nominaux susceptibles de dénoter une situation (accomplissement, achèvement) **et/ou** un objet (Pustejovsky, 1995 ; Godard et Jayez, 1996 ; Milićević & Polguère, 2010).

– *Son **exposé** a été très long et très ennuyeux.* (ACC•OBJETinfo)

2.2.3 Hiérarchies (verbale et nominale) des classes aspectuelles simples

Nous récapitulons, dans les figures 1 et 2, les classes aspectuelles qui nous ont servi à la description, présentées ici sous forme de hiérarchie.

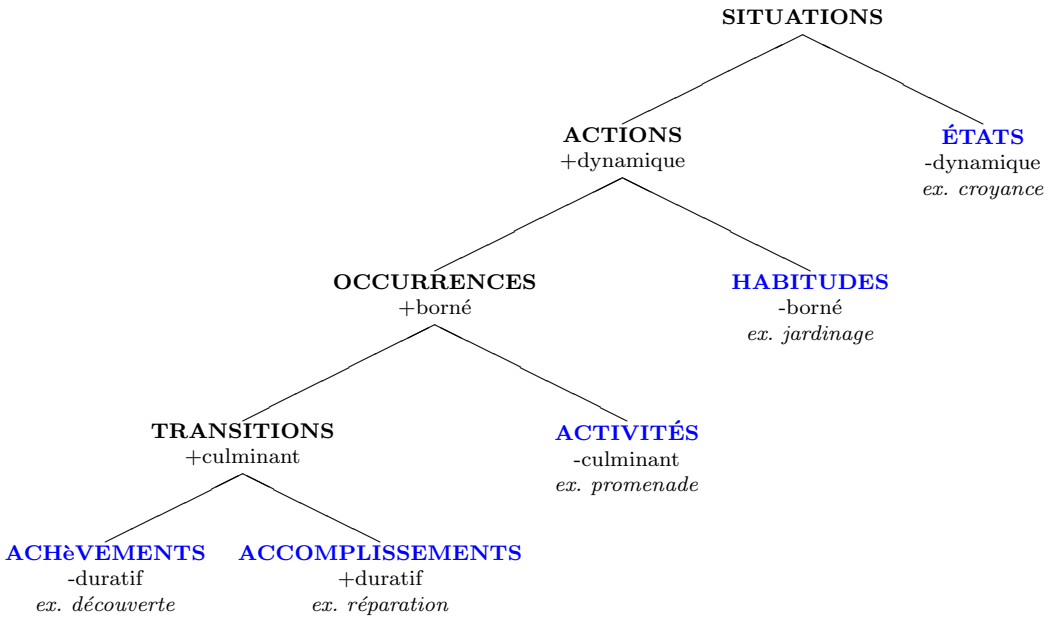


FIG. 2 – Hiérarchie des classes aspectuelles simples pour les noms

3 Le corpus Nomage

3.1 Présentation du corpus

Le corpus utilisé pour l'annotation des nominalisations est le French TreeBank, corpus du français développé il y a une dizaine d'années au laboratoire LLF (CNRS & Université Paris 7) sous la direction d'Anne Abeillé. Le French TreeBank regroupe des articles du journal *Le Monde* parus entre 1989 et 1995 (environ 1 million de mots) et propose trois principaux niveaux d'annotation linguistiques :

- annotation morpho-syntaxique (adjectif, nom commun, préposition, etc.)
- annotation en constituants (groupe nominal, groupe prépositionnel, etc.)
- annotation en fonctions (sujet, objet, etc.)

Pour l'heure, environ 11 800 nominalisations ont été extraites de la sous-partie du corpus annotée en fonctions (soit à peu près la moitié du French TreeBank). Un certain nombre de ces nominalisations n'ont toutefois pas été analysées dans le cadre du projet Nomage, pour l'une des trois raisons suivantes :

- la nominalisation est un nom dérivé d'un adjectif (par ex. *indulgence*, dérivé de *indulgent*) or nous n'étudions ici que les noms dérivés de verbes ;
- le nom est lié à un verbe mais le sens de la dérivation n'apparaît pas clairement. C'est le cas, typiquement, des conversifs. Par exemple, on ne sait pas si *voyager* est

dérivé de *voyage* ou si c'est l'inverse ;
- il ne s'agit pas d'une nominalisation. Les nominalisations sont extraites automatiquement en fonction d'un suffixe (par exemple *-tion*, *-age*, etc.) et d'une longueur minimale de caractères précédant ce suffixe. Cette heuristique nous conduit immanquablement à la sélection de candidats qui ne sont pas des nominalisations (par ex. *sarcophage*).

Une "stop-list" a été dressée pour exclure automatiquement la plupart des cas mentionnés ci-dessus. Les annotateurs se sont chargés d'identifier les erreurs restantes.

3.2 Annotation aspectuelle des nominalisations en corpus

L'originalité du projet Nomage repose en partie sur l'application de tests sur des exemples non construits. Ces tests sont censés mettre en évidence certaines propriétés sémantiques du candidats, notamment ses propriétés aspectuelles.

3.2.1 Présentation des 10 tests

Nous donnons ici une description rapide des 10 appliqués à chaque occurrence d'une nominalisation dans le corpus (cf. annexe 1 pour une description plus complète des tests).

T1_Plusieurs Ce test consiste à remplacer le déterminant du nom par "plusieurs" ou à ajouter cette séquence, si le nom n'a pas de déterminant. Sauf blocage dû au contexte de la phrase, ce test doit rendre compte du **caractère comptable ou massif** de l'unité testée.

T2_Avoir lieu Dans ce test, le nom est modifié au moyen d'une relative de la forme "qui AVOIR lieu + complément de temps". Sauf blocage dû au contexte de la phrase, ce test doit révéler si l'unité testée dénote ou non un **événement** (autrement dit une situation dynamique bornée).

T3_Éprouver/ressentir Dans ce test, le nom est modifié au moyen d'une relative de la forme "que x EPROUVER/RESSENTIR (+ complément de temps)". Sauf blocage dû au contexte de la phrase, ce test doit révéler si l'unité testée dénote ou non un **état**, en l'occurrence un état psychologique.

T4_Un peu de Ce test consiste à remplacer le déterminant du nom par la séquence "un peu de" ou à ajouter cette séquence, si le nom n'a pas de déterminant. Sauf blocage dû au contexte de la phrase, ce test doit rendre compte du **caractère comptable ou massif** de l'unité testée.

T5_Durer x temps Dans ce test, le nom est modifié au moyen d'une relative de la forme "qui DURER x temps". Sauf blocage dû au contexte de la phrase, ce test doit révéler si l'unité testée dénote ou non un **événement duratif**.

T6_Se trouver Dans ce test, le nom est modifié au moyen d'une relative de la forme "qui SE TROUVER (+ complément de lieu)". Sauf blocage dû au contexte de la phrase, ce test doit révéler si l'unité testée dénote ou non un **objet**, en l'occurrence un objet concret.

T7_Effectuer/procéder Dans ce test, le nom est modifié au moyen d'une relative de la forme "que x ÉFFECTUER (+ complément de temps)" et/ou d'une relative de la forme "auquel on PROCÉDER (+ complément de temps)". Sauf blocage dû au contexte de la phrase, ce test doit révéler si l'unité testée dénote ou non un **événement**.

T8_État de Ce test consiste à placer la séquence "état de" juste à la gauche du candidat, c'est-à-dire entre le candidat et son déterminant. Sauf blocage dû au contexte de la phrase, ce test doit révéler si l'unité testée dénote ou non un **état**.

T9_Se dérouler Dans ce test, le nom est modifié au moyen d'une relative de la forme "qui SE DÉROULER (+ complément de temps)". Sauf blocage dû au contexte de la phrase, ce test doit révéler si l'unité testée dénote ou non un **événement duratif**.

T10_Cardinal Le test consiste à remplacer le déterminant (qu'il soit défini ou indéfini) par un cardinal (par exemple, *trois*, *trente*, *deux cents*, etc.). Sauf blocage dû au contexte de la phrase, ce test doit rendre compte du **caractère comptable ou massif** de l'unité testée.

3.2.2 Attribution d'une classe aspectuelle à partir du résultat des tests

T1 plrs	**T2** avoir lieu	**T3** éprou-ver	**T4** un peu	**T5** durer x tps	**T6** se trouver	**T7** pro-céder	**T8** état de	**T9** se dérouler	**T10** card.	**classe aspectuelle**
-	○	○	-	-	•/-	○	○	○	-	OBJET
-	○	•/-	-	-	○	○	•/-	○	-	ETAT
○	○	○	•	-	○	○	○	○	○	HAB
•/-	•	○	○	○	○	-	○	○	•	EvtPonctuel
•/-	•	○	○	•/-	○	-	○	•/-	•/-	EvtDuratif ou EvtPonctuel•ETAT

3.2.3 Exemple d'annotation de noms déverbaux

Nous présentons ci-dessous un exemple de fiche annotée (figure 3). La fiche débute par une phrase extraite du French Treebank dans laquelle figure le candidat à annoter (indiqué en gras) et par l'indication du GN dans lequel figure ce candidat. Puis viennent la partie consacrée aux propriétés observées (les propriétés déjà renseignées sont ici soulignées) et la partie consacrée aux propriétés inférées. La dernière ligne de la fiche contient la classe aspectuelle inférée des résultats de l'application des tests.

*Ça, c'est colossal, parce qu'enfin, jusqu'à l'**annexion**, les pays baltes et scandinaves,question niveau de vie, c'était du pareil au même.*
*l'**annexion***
wordForm : annexion **isLemma :** annexion **morphoCue :** xion **hasMorpho :** N-C-fs **isDerivedFromVerb :** annexer **isSyntHead :** yes
T1_Plusieurs : no **T2_Avoir lieu :** yes → l'annexion qui a eu lieu l'année dernière **T3_Éprouver/ressentir :** no **T4_Un peu de :** no **T5_Durer x temps :** yes → l'annexion, qui a duré 3 ans **T6_Se trouver :** no **T7_Éffectuer/procéder :** yes → l'annexion à laquelle ont procédé les russes **T8_État de :** no **T9_Se dérouler :** no **T10_Card :** no
Classe aspectuelle inférée : EvtDuratif ou EvtPonctuel•ETAT

FIG. 3 – Annotation d'une occurrence du nom *annexion*

3.3 Réalisation syntaxique des arguments dans le corpus

La réalisation syntaxique des arguments sémantiques des noms prédicatifs[6] est, comme nous l'avons dit précédemment, décrite au niveau du corpus (autrement dit au niveau des occurrences des nominalisations) et non au niveau du lexique (voir *supra*.

L'encodage de la réalisation syntaxique des arguments s'inspirent des descriptions produites dans le cadre de la LEC (Mel'čuk *et al.*, 1995), à quelques différences près : il nous semble par exemple intéressant de distinguer, dans le cadre d'une étude de l'aspect, les groupes nominaux incluant un déterminant (Gdét) de ceux qui n'en inclut pas (GN). Voici quelques exemple d'encodage de la réalisation syntaxique des arguments de noms déverbaux apparaissant dans le corpus :

- Occurrences du lexème "ANNULATION de Y par X"

  → *Elle exige l'**annulation** de ces délibérations et a décidé d'organiser un sit-in quotidien, à l'heure du repas, à partir du 15 janvier.*

  Réalisation syntaxique : X=∅, Y=de Gdet

  → *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.*

  Réalisation syntaxique : X=par Gdét, Y=de Gdét

- 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.*

  Réalisation syntaxique : X=adj. rel., Y=de GN

- 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.*

  Réalisation syntaxique : X=∅, Y=de Gdét, Vsup=X assurer det N

Comme le montrent ces exemples, seuls sont encodés les arguments réalisés en tant que **dépendants syntaxiques** de la nominalisation. Lorsqu'un argument est présent dans la phrase mais lié via un verbe support à la nominalisation, il est encodé au niveau de l'attribut VSup, comme illustré ci-dessus avec COMMERCIALISATION : la 'personne qui

[6] La réalisation syntaxique des arguments sémantiques de leurs **verbes sources** n'est pas décrite dans le cadre du projet Nomage. Le lexique DicoValence, duquel nous nous sommes largement inspirés pour décrire la structure argumentale sémantique des verbes, propose quand à lui une telle description (van den Eynde et Mertens, 2003).

commercialise' (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. Un compte rendu complet de l'encodage des arguments exprimés est présenté dans l'annexe 2, page 27

4 Le lexique Nomage

Le lexique décrit l'ensemble des lexèmes nominaux instanciés dans le corpus ainsi que leur source verbale. La description consiste pour l'essentiel à identifier leur structure argumentale complète (section 4.3) à leur attribuer une classe aspectuelle (section 4.4).

4.1 Présentation générale : une entrée du lexique Nomage

Unités de base du lexique, les **lexèmes** sont caractérisés par neuf attributs, illustré dans la figure 4 avec le lexème ASSOCIATION#1 :

ID = `42`
vocable = `association`
num_acc = `1`
pdd = `nom`
exemple = `L'association du canadien Northern Telecom avec Matra dans les télécommunications passe par une offre publique d'achat simplifiée.`
struct_arg_sem = `association de X avec Y`
classe_aspectuelle = `ACHs`
lex_source = `44`
occurrences = {4439, 725}

FIG. 4 – L'entrée `association#1` dans le lexique Nomage

1. leur **identifiant**, unique pour chaque lexème,
2. leur **vocable**, autrement dit leur lemme,
3. leur **numéro d'acception**, utile quand le vocable est polysémique,
4. leur **partie du discours** (nom ou verbe),
5. un **exemple** illustratif choisi parmi l'ensemble de ses occurrences dans le corpus du French Treebank,
6. leur **structure argumentale** sémantique, c'est-à-dire la formulation de leurs arguments, si le lexème décrit est un prédicat,
7. la **classe aspectuelle** à laquelle appartient le lexème décrit,
8. le **lexème source** dont est dérivé le lexème nominal. Dans l'exemple donné ci-dessous, la source verbale de `association#1` est le lexème ayant pour identifiant le numéro 44,

9. les **occurrences** de ce lexème dans le corpus. La valeur de cet attribut est une liste d'identifiants de phrases du corpus dans lesquelles le lexème a été annoté. Dans notre exemple, le lexème ASSOCIATION#1 n'apparaît que dans les phrases numérotées 4439 et 725. Les lexèmes verbaux n'ont quant à eux pas d'occurrences identifiées dans le corpus.

4.2 Quelques chiffres

Le tableau ci-dessous présente quelques données chiffrées du lexique :

lexèmes nominaux	737
vocables nominaux	656
degré de polysémie nominale	1,12
lexèmes verbaux	679
vocables verbaux	648
degré de polysémie verbale	1,04
lexèmes nominaux dénotant des OBJET	82
lexèmes nominaux dénotant des ETAT	64
lexèmes nominaux dénotant des EVT duratifs	279
lexèmes nominaux dénotant des EVT ponctuels	128
lexèmes nominaux ayant une classe complexe (ACC•OBJET, ACH•ETAT, ...)	184

• Remarques sur la polysémie

Le faible degré de polysémie (1,12 lexèmes par vocable) s'explique sans doute essentiellement par le fait que le corpus utilisé est relativement restreint (500 000 mots) et relativement spécialisé (corpus journalistique).

La polysémie des noms dérivés de verbes a deux sources principales :

1. **Polysémie héritée du verbe source**. Dans notre exemple, PROMOTION#1 est hérité du verbe "la personne X PROMOUVOIR#1 l'individu Y au poste Z" tandis que PROMOTION#2 est hérité du verbe "la personne X PROMOUVOIR#2 Y".
   - PROMOTION#1 de Y à Z accordée par X
     C'est arrivé après sa promotion au poste de directeur financier.
   - PROMOTION#2 de Y par X
     Chirac va faire la promotion de son livre en plein marasme judiciaire.

2. **Polysémie propre au nom** : relation de métonymie entre une action et un des participants de l'action, dans notre exemple, la chose installée
   - INSTALLATION#1 de Y par X
     L'annonce de l'installation du poulailler géant suscite une vigoureuse protestation des syndicats d'exploitants agricoles.
   - INSTALLATION#2
     La vétusté des installations y est pour beaucoup.

4.3 Structure argumentale des prédicats décrits

Le lexique Nomage décrit systématiquement les arguments sémantiques des prédicats nominaux et verbaux, c'est-à-dire les différents participants requis pour définir la situation dénotée par un prédicat donné. Les arguments sémantiques sont représentées par des variables (X, Y, Z)[7] , comme illustré ci-dessous :

- ADHÉSION#1 de X à Y [*son adhésion au parti socialiste*]
- EXPLOITATION#1 de Y par X [*l'exploitation de l'homme par l'homme*]
- X RESTITUER Y à Z [*On lui a restitué 400 hectares de forêt.*]

Les arguments sémantiques d'une nominalisation correspondent généralement à ceux du lexème verbal dont est dérivé cette nominalisation, comme illustré ci-dessous. Dans les deux cas, 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').

- X EXPÉDIER Y à Z
  On a fabriqué l'extracteur pilote et on l'a expédié au Mozambique en novembre 1992.
- EXPÉDITION de Y à Z par X
  Rien de sérieux n'a été fait pour faciliter les expéditions de produits élaborés véritablement agricoles (et non agro-alimentaires).

Il arrive cependant que les différents participants de la situation dénotée par le prédicat 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 agacée' est représentée par Y dans l'entrée verbale et par X dans l'entrée nominale. Inversement, 'ce qui agace' est représenté par la variable X pour le verbe et par

[7] Ces variables devront, à terme, être typées :
- ADHÉSION#1 de la *personne* X à l'*organisation* Y
- EXPLOITATION#1 de Y par la *personne* X
- la *personne* X RESTITUER la *chose* Y à la *personne* Z

la variable Y pour le nom.

- X AGACER Y
  Sa façon de parler m'agace.
- AGACEMENT de X dû à Y
  Mais, en dénonçant avec un certain agacement la "french mafia", nos confrères étrangers ne rendent-ils pas indirectement hommage à une nouvelle forme d'influence française ?

4.4 Classe aspectuelle d'un lexème

Une fois décrite la structure argumentale des lexèmes, nous leur attribuons une classe aspectuelle. Les tests utilisés pour cette étape de la description sont décrits dans cette section.

4.4.1 Rappel des étiquettes de classes aspectuelles utilisées

classes aspectuelles pour les verbes	ETAT, ACT, ACC, ACH, ACH•ETAT, ACC•ETAT, ACC-ACT
classes aspectuelles pour les noms	les mêmes + HAB, OBJET(Info/psy)

4.4.2 Tests pour l'attribution de classes aspectuelles aux verbes

Verbes dynamiques vs. statifs – Les premiers tests appliqués pour attribuer une classe aspectuelle aux lexèmes verbaux reposent sur l'opposition stativité/dynamicité. Les lexèmes verbaux qui passent les tests 1, 2 et 3 sont dynamiques. Ceux qui ne passent pas ces tests sont des **états** (ETAT).

1) *X être en train de V*
X RÉPARER Y : *Pierre est en train de réparer sa moto.* ⇒ verbe dynamique
X POSSÉDER Y : **Pierre est en train de posséder plusieurs maisons.* ⇒ ETAT

2) *Alpha voir X V*
X CONSTITUER Y : *Je le vois constituer son équipe* ⇒ verbe dynamique
X CONNAÎTRE Y : **Je le vois connaître la réponse* ⇒ ETAT

3) *ce qu'a fait X, c'est V*
X INSTALLER Y : *Ce qu'a fait fait Pierre, c'est installer le poulailler.* ⇒ verbe dynamique
X ADMIRER Y : **Ce qu'a fait fait Pierre, c'est admirer ses grand-parents.* ⇒ ETAT

Verbes dynamiques téliques vs. atéliques – Nous distinguons ensuite, parmi les verbes dynamiques, les verbes téliques des verbes atéliques, et, parmi les verbes téliques,

ceux qui sont duratifs de ceux qui ne le sont pas. Nous utilisons pour cela les tests bien connus dans la littérature. Soit une phrase simple P constituée du verbe à tester mis au passé composé, d'un sujet délimité SQA et d'un objet délimité SQA[8] :

4) *P en x temps* : les verbes qui passent ce test sont soit des **accomplissements** (ACC) – *en x temps* porte alors sur l'ensemble de la situation dynamique dénotée par le verbe –, soit des **achèvements** (ACH) – *en x temps* porte alors sur la phase préparatoire de la situation dénotée par le verbe (équivalent alors à *au bout de x temps*)[9]. Les verbes qui ne passent pas ce test sont atéliques.

X RÉPARER Y : *Pierre a réparé sa moto en 1 heure.* ⇒ ACC

X ACQUÉRIR Y : *Il a acquis cette maison en deux jours.* ⇒ ACH

X ADMINISTRER Y : **Il a administré ses biens en qualques années.* ⇒ ACT

5) *P pendant x temps* : les verbes qui passent ce test sont soit des **activités** (ACT) – *pendant x temps* porte alors sur l'ensemble de la situation dynamique dénotée par le verbe –, soit des **achèvements-statifs** ou des **accomplissements-statifs** – *pendant x temps* porte alors sur la situation stative formant une sous-partie de la situation dénotée par le verbe.

X MANIFESTER : *Ils ont manifesté pendant plusieurs heures.* ⇒ ACT

X EMPRISONNER Y : *Ils l'ont emprisonné pendant des années.* ⇒ ACHs

X ENVAHIR Y : *Ils ont envahi le pays pendant des années.* ⇒ ACCs

6) le test du **paradoxe imperfectif** (Dowty, 1979).

X RÉPARER Y : *Ils réparaient le toit mais ont été interrompu.* **n'implique pas** *Ils ont réparé le toit.* ⇒ ACC

X MANIFESTER Y : *Ils manifestaient mais on été interrompu.* **implique** *Ils ont manifesté.* ⇒ ACT

Verbes téliques duratifs vs. ponctuels – Plusieurs tests nous permettent de distinguer, parmi les verbes téliques, ceux qui sont duratifs (les accomplissements, ACC), de ceux qui ne le sont pas (les achèvements, ACH)

7) *finir/terminer de V*

X INSTALLER Y : *Il a fini d'installer la fenêtre.* ⇒ ACC

[8] On pourrait généraliser à : tout argument du verbe doit être "délimité SQA". Tous les lexèmes verbaux décrits dans le lexique Nomage n'ont, en effet, pas toujours exactement deux arguments correspondant respectivement au sujet syntaxique et à l'object syntaxique. En voici quelques exemples :

X SE DÉROULER Y [*La séance se déroule normalement.*]

X ALIMENTER Y en Z [*La rivière alimente le moulin en eau.*]

X CANDIDATER à Y [*Il a candidaté à l'appel à projet.*]

[9] Notons toutefois que tous les achèvement ne sont pas compatibles avec *en x temps*.

X SIGNER Y : * *Il a fini de signer la feuille.* ⇒ ACH

4.4.3 Tests pour l'attribution de classes aspectuelles aux noms

Noms occurrentiels vs. non-occurrentiels Les deux premiers tests appliqués distinguent les noms qui dénotent une occurrence de ceux qui ne dénotent pas une occurrence (les états et les habitudes). Le test 1 est appliqué en premier car il nous semble être le plus déterminant : on conclut qu'il ne s'agit pas d'une occurrence si le nom testé n'accepte pas "avoir lieu". On ne déduit rien en revanche des lexèmes nominaux ne passant pas ces deux tests.

1) *N avoir lieu*

MANIFESTATION de X pour/contre Y : *La manifestation a eu lieu hier.* ⇒ occ

RÉPARATION de Y par X : *La réparation de la moto a eu lieu hier.* ⇒ occ

2) *VSup dynamiques (effectuer, réaliser, procéder à, accomplir, ...) N*

SAUVETAGE de Y par X : *Il a effectué le sauvetage d'un randonneur.* ⇒ occ

DÉCOUPAGE de Y par X : *Qui veut procéder au découpage du poulet ?* ⇒ occ

Noms non-occurrentiels de type habitude vs. état Si le nom testé n'est pas une occurrence, il faut savoir s'il correspond à un **état** (ETAT) ou à une **habitude** (HAB). Les cinq tests suivants nous permettent, dans une certaine mesure, de le déterminer. Nous ne déduisons rien des lexèmes qui ne passent pas ces tests.

3) *faire du N*

JARDINAGE effectué par X : *faire du jardinage.* ⇒ HAB

VULGARISATION de Y par X : *faire de la vulgarisation.* ⇒ HAB

4) *un état de N*

SOUFFRANCE de X à cause Y : *un état de souffrance* ⇒ ETAT

DÉPENDANCE de X à Y : *un état de dépendance* ⇒ ETAT

5) *éprouver, ressentir du/de la N*

ADMIRATION de X pour Y : *éprouver de l'admiration* ⇒ ETAT

FRUSTRATION de X à cause de Y : *ressentir de la frustration* ⇒ ETAT

6) *faire preuve de N*

DÉTERMINATION de X à Y : *faire preuve de détermination* ⇒ ETAT

IMAGINATION de X : *faire preuve d'imagination* ⇒ ETAT

7) *agir avec N*

ENTÊTEMENT de X : *agir avec entêtement* ⇒ ETAT

DÉFIANCE de X vis-à-vis de Y : *agir avec défiance.* ⇒ ETAT

Noms occurrentiels duratifs vs. ponctuels Si au contraire les tests 1 et 2 indiquent que le nom testé est une occurrence, il faut savoir si cette occurrence est durative (accomplissement ou activité) ou non, auquel cas il s'agit d'un **achèvement** (ACH). Les quatre tests suivants permettent de le déterminer. La combinaison des tests 8, 9 et 11 permet par ailleurs d'identifier les **achèvements-statifs** (ACHs).

8) *commencer/continuer le N*

FABRICATION de Y par X : *commencer la fabrication* ⇒ ACC

DISCUSSION entre X et Y : *commencer la discussion* ⇒ ACT

DÉMISSION de X du poste Y : **continuer la démission* ⇒ ACH

9) *N se dérouler en/pendant x temps*

DÉMOLITION de Y par X : *La démolition s'est déroulé en 2 heures.* ⇒ ACC

MANIFESTATION de X pour/contre Y : *La manifestation s'est déroulé pendant deux heures.* ⇒ ACT

ARRIVÉE de X à Y : **son arrivée s'est déroulée en/pendant x temps* ⇒ ACH

10) *en cours de N, N en cours*

CRÉATION de Y par X : *la création en cours* ⇒ ACC

EXPLOITATION de Y par X : *l'exploitation en cours* ⇒ ACT

11) *N dure x temps*

RÉPARATION de Y par X : *La réparation dure 3 heures.* ⇒ accomplissement ou activité[10]

Noms occurrentiels téliques vs. atéliques Enfin, les quatre tests suivants permettent de distinguer les **activités** (ACT) des **accomplissements** (ACC). En outre la combinaison des tests 12 et 14 nous permettent d'identifier les **activité-accomplissement** (ACT-ACC) : un nom qui passe le test du paradoxe imperfectif et qui accepte d'être modifié par "en x temps" est de type ACT-ACC.

12) *paradoxe imperfectif*

MANIFESTATION de X pour/contre Y : *La manifestation a été interrompue.* **implique** *Il y a eu manifestation.* ⇒ ACT

RÉPARATION de Y par X : *La réparation a été interrompue.* **n'implique pas** *Il y a eu réparation.* ⇒ ACC

13) *le N a été mené à bien hier*

DÉMEMBREMENT de Y par X : *Le démembrement a été mené à bien hier.* ⇒ ACC

14) *en voie de N*

PRIVATISATION de Y par X : *en voie de privatisation* ⇒ ACC

[10]Si durer porte sur l'action et non sur la situation résultante.

Annexe 1 : détail des tests appliqués en corpus

Chacun des tests présentés ci-dessous est illustré de deux exemples, un premier exemple où le test marche (l'annotateur doit alors indiquer **yes**), un second exemple où le test ne marche pas (**no**). Dans chaque cas, le nom candidat est souligné (avec son éventuel déterminant) tandis que la modification correspondant à l'application du test est indiquée en gras.

T1_Plusieurs Ce test consiste à remplacer le déterminant du nom par "plusieurs" ou à ajouter cette séquence, si le nom n'a pas de déterminant.

- *Une allocation sera versée en décembre.* → ***Plusieurs** allocations seront versées en décembre.* (**yes**)
- *Tout au plus des petites choses à changer sur l'intégration.* → **Tout au plus des petites choses à changer sur **plusieurs** intégrations.* (**no**)

*****

T2_Avoir lieu Dans ce test, le nom est modifié au moyen d'une relative de la forme "qui AVOIR lieu + complément de temps".

- *La décentralisation des universités apparaît comme l'un des grands débats de l'année.* → *La décentralisation des universités **qui a lieu en ce moment** apparaît comme l'un des grands débats de l'année.* (**yes**)
- *Entre gens qui ont des vraies convictions, il peut y avoir convergences.* → **Entre gens qui ont des vraies convictions, il peut y avoir convergences **qui ont lieu en ce moment**.* (**no**)

Remarque sur le verbe de la relative Le verbe de la relative (indiqué en lettres capitales) doit être conjugué à un temps (passé, présent, futur, etc.) et à un mode (conditionnel, indicatif) qui conviennent dans le contexte de la phrase.

Remarque sur le complément de temps Ce complément peut être de l'une des formes suivantes :

- il peut s'agir d'un nom (*lundi*), d'un groupe nominal (*le 23 février, ce mois-ci*) ou d'un adverbe (*hier, demain*).
- il peut s'agir d'un groupe prépositionnel (*à 8h, en février dernier, au printemps, en 1987...*). On peut également, **au besoin**, recourir à un complément de durée (*pendant deux jours*) ou à un complément de fréquence (*chaque samedi*).

Remarque sur le type de la relative Les propositions relatives peuvent être de deux types selon la façon dont elles modifient le nom auxquelles elles s'appliquent :

spécificatives, comme illustré en (1.a) ou **explicatives**, comme illustrées en (1.b). Cette différence se marque dans la phrase par l'utilisation ou non de virgules encadrant la relative.

(1) a. *Les randonneurs **qui sont fatigués** pourront faire une pause.*

b. *Les randonneurs**, qui sont fatigués,** pourront faire une pause.*

Les annotateurs doivent, chaque fois que cela est possible, utiliser une relative spécificative. Si le contexte rend cela impossible ou bizarre, il est possible d'essayer d'appliquer le test avec une relative explicative.

Remarque sur la position de la relative La relative se place, lorsque cela est possible, le plus à droite possible du nom candidat. Toutefois, il arrive fréquemment que la relative ne puisse être placée qu'après le(s) complément(s) prépositionnel(s) du nom, comme c'est le cas ci-dessus avec *la décentralisation des universités.*

Notons que si le nom candidat est déjà modifié par une relative, il convient de coordonner cette relative avec la relative du test au moyen d'un *ou*, d'un *et* ou d'un *mais.*

- *C'est une manifestation sans ambition spatiale excessive qui s'est installée dans les salles restantes. → C'est une manifestation sans ambition spatiale excessive **qui a lieu le 4 mars** et qui s'est installée dans les salles restantes* (**yes**)

*****

T3_Éprouver/ressentir Dans ce test, le nom est modifié au moyen d'une relative de la forme "que x EPROUVER/RESSENTIR (+ complément de temps)".

- *L'admiration, sinon la confiance, se sont émoussées. → L'admiration, sinon la confiance, **qu'on éprouvait depuis longtemps** se sont émoussées.* (**yes**)
- *Cette revendication vient d'être rappelée par l'association nationale des élus locaux. → *Cette revendication**, qu'on éprouve depuis longtemps,** vient d'être rappelée par l'association nationale des élus locaux.* (**no**)

Remarque sur le verbe de la relative Cf. T2

Remarque sur le complément de temps Cf. T2

Remarque sur le type de la relative Cf. T2

Remarque sur la position de la relative Cf. T2

Remarque sur l'application du test Le test s'applique si l'une et/ou l'autre des relatives (celle avec *éprouver* ou celle avec *ressentir*) s'applique. Il ne s'applique pas si aucune des deux relatives n'est possible.

*****

T4_Un peu de Ce test consiste à remplacer le déterminant du nom par la séquence "un peu de" ou à ajouter cette séquence, si le nom n'a pas de déterminant.

- *Mais, en dénonçant avec un certain agacement la "french mafia", nos confrères étrangers ne rendent-ils pas indirectement hommage à une nouvelle forme d'influence française ? → Mais, en dénonçant avec **un peu d'**agacement la "french mafia", nos confrères étrangers ne rendent-ils pas indirectement hommage à une nouvelle forme d'influence française ?* (**yes**)
- *Une réunion du courant Socialisme et République samedi soir tranchait la question : l' hostilité à "l'inconstance politique" de Mr Dray l' emportait. → *Une réunion du courant Socialisme et République samedi soir tranchait **un peu de** question : l' hostilité à "l'inconstance politique" de Mr Dray l' emportait.* (**no**)

Remarque sur les noms au pluriel Les noms au pluriel doivent pouvoir être mis au singulier :

- *Un peu de convergences* : *un peu de convergence* (**yes**)
- *Un peu de logements* : **un peu de logement* (**no**)

Remarque sur un éventuel changement de sens Il se peut la transformation change le sens du candidat, comme illustré dans la phrase ci-dessous. On considérera dans ces cas-là que le test s'applique, sachant que l'analyse post-annotation reviendra sur cette différence sémantique.

- *L'élue de l'Essonne ne cachait pas que l'expérience la tentait. → L'élue de l'Esonne ne cachait pas qu'**un peu d'**expérience la tentait.* (**yes**)

*****

T5_Durer x temps Dans ce test, le nom est modifié au moyen d'une relative de la forme "qui DURER x temps".

- *La décentralisation des universités apparaît comme l'un des grands débats de l'année. → La décentralisation des universités**, qui durera deux ans,** apparaît comme l'un des grands débats de l'année.*
- *Monsieur Jospin a insisté sur la part que doivent prendre les collectivités locales dans les décisions et les investissements. → *Monsieur Jospin a insisté sur la part que doivent prendre les collectivités locales dans les décisions **qui dureront trois jours** et les investissements.* (**no**)

Remarque sur le complément de durée Dans le complément de durée, *x* doit correspondre à un chiffre, *temps* à minute, heure, jour, semaine, mois ou année.

Remarque sur le verbe de la relative Cf. T2

Remarque sur le type de la relative Cf. T2
Remarque sur la position de la relative Cf. T2

*****

T6_Se trouver Dans ce test, le nom est modifié au moyen d'une relative de la forme "qui SE TROUVER (+ complément de lieu)".

- *Le gouvernement a décidé d'accorder la maîtrise d'ouvrage aux collectivités locales, pour les constructions universitaires.* → *Le gouvernement a décidé d'accorder la maîtrise d'ouvrage aux collectivités locales, pour les constructions universitaires* ***qui se trouveront en banlieue****.* (**yes**)
- *Mais de leur côté, les collectivités demandent que cette participation s'accompagne d'une extension de leurs compétences à l'enseignement supérieur.* → **Mais de leur côté, les collectivités demandent que cette participation****, qui se trouvera en banlieue,** *s'accompagne d'une extension de leurs compétences à l'enseignement supérieur.* (**no**)

Remarque sur le type de la relative Cf. T2
Remarque sur la position de la relative Cf. T2

*****

T7_Effectuer/procéder Dans ce test, le nom est modifié au moyen d'une relative de la forme "que x ÉFFECTUER (+ complément de temps)" et/ou d'une relative de la forme "auquel on PROCÉDER (+ complément de temps)".

- *ça c'est colossal, parce qu'enfin, jusqu'à l'annexion, les pays baltes et scandinaves, question niveau de vie, c'était du pareil au même* → *ça c'est colossal, parce qu'enfin, jusqu'à l'annexion* ***à laquelle on a procédé en 2000****, les pays baltes et scandinaves, question niveau de vie, c'était du pareil au même* (**yes**)
- *Entre gens qui ont des vraies convictions, il peut y avoir convergences.* → **Entre gens qui ont des vraies convictions, il peut y avoir convergences****, auxquelles on procédera bientôt****.* (**no**)

Remarque sur le verbe de la relative Cf. T2
Remarque sur le complément de temps Cf. T2
Remarque sur le type de la relative Cf. T2
Remarque sur la position de la relative Cf. T2
Remarque sur l'application du test Cf. T3

*****

T8_État de Ce test consiste à placer la séquence "état de" juste à la gauche du candidat, c'est-à-dire entre le candidat et son déterminant.

- *Sa vie et la diversité de son talent en auront fait une sorte de voyageur "professionnel" dans une Europe en pleine effervescence.* → *Sa vie et la diversité de son talent en auront fait une sorte de voyageur "professionnel" dans une Europe en plein* ***état d'****effervescence.* (**yes**)
- *Il y avait un moyen simple de prouver cette intention.* → **Il y avait un moyen simple de prouver cet* ***état d'****intention.* (**no**)

*****

T9_Se dérouler Dans ce test, le nom est modifié au moyen d'une relative de la forme "qui SE DÉROULER (+ complément de temps)".

- *Chaque année ou presque, La Royal Academy of Art consacre une de ses expositions majeures à l'architecture.* → *Chaque année ou presque, La Royal Academy of Art consacre une de ses expositions majeures,* ***qui se déroule en général au printemps,*** *à l'architecture.* (**yes**)
- *Cette revendication vient d'être rappelée par l'association nationale des élus locaux.* → **Cette revendication,* ***qui se déroule aujourd'hui,*** *vient d'être rappelée par l'association nationale des élus locaux.* (**no**)

Remarque sur le verbe de la relative Cf. T2
Remarque sur le complément de temps Cf. T2
Remarque sur le type de la relative Cf. T2
Remarque sur la position de la relative Cf. T2

*****

T10_Cardinal Le test consiste à remplacer le déterminant (qu'il soit défini ou indéfini) par un cardinal (par exemple, *trois*, *trente*, *deux cents*, etc.). Il est également possible de placer le cardinal entre le candidat et son déterminant, lorsque celui-ci est défini ou encore d'ajouter le cardinal en position de déterminant lorsque le nom n'a pas déjà un déterminant.

- *Quelles sont les possibilités et les intentions des différents acteurs ?* → *Quelles sont les possibilités et les intentions des* ***trois*** *différents acteurs ?* (**yes**)
- *Mais, de leur côté, les collectivités demandent que cette participation s'accompagne d'une extension de leurs compétences à l'enseignement supérieur.* → **Mais, de leur côté, les collectivités demandent que cette participation s'accompagne de* ***trois*** *extensions de leurs compétences à l'enseignement supérieur.* (**no**)

Annexe 2 : Réalisation syntaxique des arguments

Explicitation des codifications choisies pour le codage de la réalisation syntaxique de la structure argumentale de surface des noms déverbaux de *Nomage*

En ce qui concerne le codage du type de complément du nom déverbal, on distingue les compléments avec article (ex : *la construction d'un logement*) des compléments sans article (ex : *la construction de logement*), d'où la distinction, qui s'appuie sur les constituants de la grammaire traditionnelle, entre :
- Groupe Déterminant (Gdét), qui implique la présence d'un article (défini ou indéfini, singulier ou pluriel) ;
- Groupe Nominal (GN) au singulier ou au pluriel, sans article apparent dans la réalisation syntaxique de surface.

Cela permet d'établir ainsi une distinction syntaxique entre :
- « la construction du logement », où « du logement », mis pour « de + le logement », = « de Gdét (défini au singulier) » ;
- « la construction des logements », où « des logements », mis pour « de + les logements » = « de Gdét (défini au pluriel) » ;
- « la construction d'une maison », où « d'une maison » = « de Gdét (indéfini au singulier) » ;
- « la construction de logements », où « de logements », mis dans la réalisation de surface pour « de + des logements » dans la structure syntaxique profonde, = « de GN (au pluriel) » ;
- « la construction de logement », où « de logement » = « de GN (au singulier) ».

En outre, il a fallu prendre en compte la codification syntaxique des arguments « cachés », tels que :
- l'adjectif possessif, abrégé en « adj poss » ;
- l'adjectif relationnel, abrégé en « adj rel ».

Ainsi, a été élaborée la codification suivante, applicable à toutes les nominalisations de du corpus et du lexique *Nomage* :
- « N de nom » quand le nom déverbal est en position de complément du nom d'un autre nom N dont il dépend (ex. : « les dépenses de construction » = « N de construction ») ;
- « N de nom de X / Y / Z » quand le nom déverbal est en position de complément du nom d'un autre nom N et qu'il est suivi de l'argument X, Y ou Z (ex. : « société de construction de moteurs d'aviation » = « N de construction de Y (= GN (au pluriel)) ») ;
- « Ø » s'il n'y a pas d'argument : ainsi, pour la séquence « cette autorisation », où le nom *autorisation* est employé seul, on a comme codification :
    - « X = Ø » ;
    - « Y = Ø » ;
    - « Z = Ø » ;
- « nom propre », quand X / Y / Z apparaissent sous la forme d'un nom propre :
    - pour la séquence « la construction d'Abuja », on aura la codification :
        - « X = Ø » ;
        - « Y = de nom propre (d'Abuja) » ;
    - pour la séquence « l'intervention de M. Modrow », on aura la codification :
        - « X = de nom propre (de M. Modrow) » ;
        - « Y = Ø » ;

- « GN », quand les arguments X / Y / Z apparaissent en réalisation syntaxique de surface sous la forme d'un Groupe Nominal sans article, au singulier ou au pluriel, ou toujours sans article mais précédé d'un adjectif qualificatif, dans la mesure où l'adjectif peut être postposé :

- pour la séquence « les constructions de logement », on a :

- « X = Ø » ;

- « Y = de GN (de logement) » ;

- pour la séquence « la construction de logements », où « de logements » en surface est mis pour « de + des logements » en profondeur, on a :

- « X = Ø » ;

- « Y = de GN (de logements) » ;

- pour la séquence « la construction de nouveaux logements », qui équivaut à « la construction de logements nouveaux », on a :

- « X = Ø » ;

- « Y = de GN (de nouveaux logements) » ;

- pour la séquence « l'exploitation de nombreux gisements est nécessaire » qui équivaut à « l'exploitation de gisements nombreux », on a :

- « X = Ø » ;

- « Y = de GN (de nombreux gisements) » ;

- « Gdét », quand les arguments X / Y / Z apparaissent sous la forme d'un Groupe déterminant avec l'article défini ou indéfini, au singulier ou au pluriel *le*/*la*/*les*, *un*/*une*/*des* (*du*/*de la*/*des N*, *d'un*/*d'une*/*des N*), ou quand ils sont précédés d'un adjectif indéfini au pluriel (*de certains N*, *de plusieurs N*, *de quelques N, etc.*) qui ne peut en aucun cas être postposé au nom (**de N certains*, **de N plusieurs*, **de N quelques, etc.*), ou quand ils sont précédés d'un numéral (*1000 N*, *10 000 N*) :

- pour la séquence « après l'approbation officielle par les actionnaires de la création du groupe Pinault – Printemps », on a la codification :

- « X = par Gdét (par les actionnaires) » ;

- « Y = de Gdét (du groupe Pinault – Printemps) » ;

- pour la séquence « l'application du permis à points par les autorités administratives sera effective dans quelques jours », on a la codification :

- « X = par Gdét (par les autorités administratives) » ;

- « Y = de Gdét (du permis à points) » ;

- pour la séquence « application d'une colle », on a :

- « X = Ø » ;

- « Y = de Gdét (d'une colle) » ;

- pour la séquence « la construction des nouveaux logements », on a :

- « X = Ø » ;

- « Y = de Gdét (des nouveaux logements) » ;

- pour la séquence « l'exploitation des nombreux gisements d'ici est rare », on a :

- « X = Ø » ;

- « Y = de Gdét (des nombreux gisements) » ;

- pour la séquence « l'exploitation des quelques gisements qui sont là est rare », on a :

- « X = Ø » ;

- « Y = de Gdét (des quelques gisements) » ;

- pour la séquence « l'exploitation de quelques gisements est nécessaire », on a :

- « X = Ø » ;

- « Y = de Gdét (de quelques gisements) » ;

- pour la séquence « l'application de certaines mesures pratiques », on a :

- « X = Ø » ;
- « Y = de Gdét (de certaines mesures pratiques) » ;

- pour la séquence « l'exploitation de plusieurs gisements est nécessaire », on a :
  - « X = Ø » ;
  - « Y = de Gdét (de plusieurs gisements) » ;
- pour la séquence « la suppression de 118 emplois dans cette entreprise », on a :
  - « X = Ø » ;
  - « Y = de Gdét (de 118 emplois) » ;
- pour la séquence « la création de 252 unités mobiles sera effective », on a :
  - « X = Ø » ;
  - « Y = de Gdét (de 252 unités mobiles) » ;

- « de num », quand on a l'expression (souvent en argument Z) d'un numéral (%, somme, monnaie, etc.) dans des phrases du type :
  - pour la séquence « une augmentation de 20% est prévue pour la fin du mois », on a :
    - « X = Ø » ;
    - « Y = Ø » ;
    - « Z = de num (de 20%) » ;
  - pour la séquence « l'augmentation de la croissance de 20% », on a :
    - « X = Ø » ;
    - « Y = de Gdét (de la croissance) » ;
    - « Z = de num (de 20%) » ;
  - pour la séquence « Le FMI prévoit l'augmentation de la dette de l'Etat de 200 millions d'euros », on a :
    - « X = Ø » ;
    - « Y = de Gdét (de la dette) » ;
    - « Z = de num (de 200 millions d'euros) » ;

- « adj poss », si X ou Y ou Z est un adjectif possessif mis pour « de X », « de Y », « de Z » :
  - pour la séquence « Vingt ans après sa création par la famille Pariente, le groupe de confection et de distribution Naf-Naf fait son entrée en Bourse », on a :
    - « X = par Gdét (par la famille Pariente) » ;
    - « Y = adj poss (sa) » ;
  - pour la séquence : « son adhésion au FMI », mis pour « l'adhésion de la Russie au FMI », on obtient :
    - « X = adj poss (son) » ;
    - « Y = à Gdét (au FMI) » ;

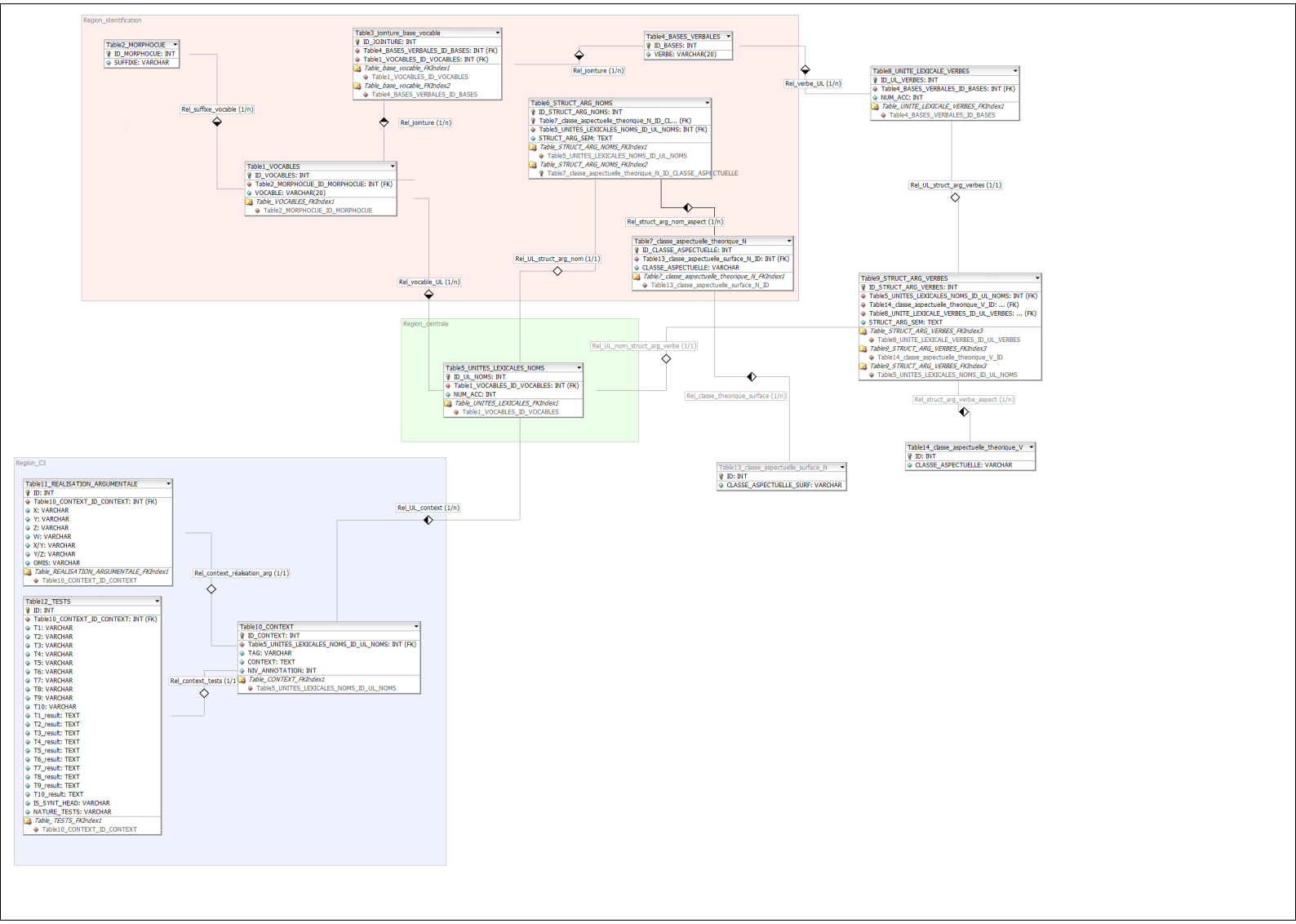
- « adj rel », si X ou Y ou Z est un adjectif relationnel mis pour « de X », « de Y », « de Z » : pour la séquence « l'adhésion russe au FMI », pour « l'adhésion de la Russie au FMI », on a :
  - « X = adj rel (russe) » ;
  - « Y = à Gdét (au FMI) » ;

- « dont » si X ou Y ou Z est le pronom relatif « dont » mis pour « de X », « de Y », « de Z » : pour la séquence « pour certains gisements dont l'exploitation est rentable », on a :
  - « X = Ø » ;
  - « Y = dont » ;

- « en » si X ou Y ou Z est le pronom adverbial « en » mis pour « de X », « de Y », « de Z » : pour la séquence « en trouver la confirmation », mis pour « trouver la confirmation de cela » :
  - « X = Ø » ;
  - « Y = en » ;

- « V inf » quand X / Y / Z apparaissent sous la forme d'un verbe à l'infinitif ou d'une proposition infinitive :
    - pour la phrase « Euronews obtient l'autorisation d'émettre », on a :
        - « X = Ø » ;
        - « Y = de V inf (d'émettre) » ;
    - pour la séquence « les hésitations de l'opposition à constituer un front commun » :
        - « X = de Gdét (de l'opposition) » ;
        - « Y = à V inf (à constituer) » ;
- « que PROP ind », quand X / Y / Z constituent une proposition à l'indicatif introduite par *que* : pour la phrase « les enquêteurs sont arrivés à la conclusion que les rabais existent » :
    - « X = Ø »;
    - « Y = que PROP ind (que les rabais existent) »;
- « que PROP subj », quand X / Y / Z constituent une proposition au subjonctif introduite par *que* : pour « les enquêteurs sont arrivés à la supposition que le suspect soit coupable » :
    - « X = Ø »;
    - « Y = que PROP subj (que le suspect soit coupable) ».

Annexe 3 : Modélisation de la base de données Nomage


Region_identification
Table2_MORPHOCUE
ID_MORPHOCUE: INT
SUFFIXE: VARCHAR
Table3_jointure_base_vocable
ID_JOINTURE: INT
Table4_BASES_VERBALES_ID_BASES: INT (FK)
Table1_VOCABLES_ID_VOCABLES: INT (FK)
Table_base_vocable_FKIndex1
Table1_VOCABLES_ID_VOCABLES
Table_base_vocable_FKIndex2
Table4_BASES_VERBALES_ID_BASES
Table4_BASES_VERBALES
ID_BASES: INT
VERBE: VARCHAR(20)
Rel_jointure (1/n)
Rel_verbe_UL (1/n)
Table8_UNITE_LEXICALE_VERBES
ID_UL_VERBES: INT
Table4_BASES_VERBALES_ID_BASES: INT (FK)
NUM_ACC: INT
Table_UNITE_LEXICALE_VERBES_FKIndex1
Table4_BASES_VERBALES_ID_BASES
Rel_suffixe_vocable (1/n)
Rel_jointure (1/n)
Table6_STRUCT_ARG_NOMS
ID_STRUCT_ARG_NOMS: INT
Table7_classe_aspectuelle_theorique_N_ID_CL... (FK)
Table5_UNITES_LEXICALES_NOMS_ID_UL_NOMS: INT (FK)
STRUCT_ARG_SEM: TEXT
Table_STRUCT_ARG_NOMS_FKIndex1
Table5_UNITES_LEXICALES_NOMS_ID_UL_NOMS
Table_STRUCT_ARG_NOMS_FKIndex2
Table7_classe_aspectuelle_theorique_N_ID_CLASSE_ASPECTUELLE
Table1_VOCABLES
ID_VOCABLES: INT
Table2_MORPHOCUE_ID_MORPHOCUE: INT (FK)
VOCABLE: VARCHAR(20)
Table_VOCABLES_FKIndex1
Table2_MORPHOCUE_ID_MORPHOCUE
Rel_UL_struct_arg_verbes (1/1)
Rel_struct_arg_nom_aspect (1/n)
Table7_classe_aspectuelle_theorique_N
ID_CLASSE_ASPECTUELLE: INT
Table13_classe_aspectuelle_surface_N_ID: INT (FK)
CLASSE_ASPECTUELLE: VARCHAR
Table7_classe_aspectuelle_theorique_N_FKIndex1
Table13_classe_aspectuelle_surface_N_ID
Rel_UL_struct_arg_nom (1/1)
Rel_vocable_UL (1/n)
Table9_STRUCT_ARG_VERBES
ID_STRUCT_ARG_VERBES: INT
Table5_UNITES_LEXICALES_NOMS_ID_UL_NOMS: INT (FK)
Table14_classe_aspectuelle_theorique_V_ID: ... (FK)
Table8_UNITE_LEXICALE_VERBES_ID_UL_VERBES: ... (FK)
STRUCT_ARG_SEM: TEXT
Table_STRUCT_ARG_VERBES_FKIndex3
Table8_UNITE_LEXICALE_VERBES_ID_UL_VERBES
Table9_STRUCT_ARG_VERBES_FKIndex3
Table14_classe_aspectuelle_theorique_V_ID
Table9_STRUCT_ARG_VERBES_FKIndex2
Table5_UNITES_LEXICALES_NOMS_ID_UL_NOMS
Region_centrale
Rel_UL_nom_struct_arg_verbe (1/1)
Table5_UNITES_LEXICALES_NOMS
ID_UL_NOMS: INT
Table1_VOCABLES_ID_VOCABLES: INT (FK)
NUM_ACC: INT
Table_UNITES_LEXICALES_FKIndex1
Table1_VOCABLES_ID_VOCABLES
Rel_classe_theorique_surface (1/n)
Rel_struct_arg_verbe_aspect (1/n)
Table14_classe_aspectuelle_theorique_V
ID: INT
CLASSE_ASPECTUELLE: VARCHAR
Table13_classe_aspectuelle_surface_N
ID: INT
CLASSE_ASPECTUELLE_SURF: VARCHAR
Region_C3
Table11_REALISATION_ARGUMENTALE
ID: INT
Table10_CONTEXT_ID_CONTEXT: INT (FK)
X: VARCHAR
Y: VARCHAR
Z: VARCHAR
W: VARCHAR
X/Y: VARCHAR
Y/Z: VARCHAR
OMIS: VARCHAR
Table_REALISATION_ARGUMENTALE_FKIndex1
Table10_CONTEXT_ID_CONTEXT
Rel_UL_context (1/n)
Rel_context_réalisation_arg (1/1)
Table12_TESTS
ID: INT
Table10_CONTEXT_ID_CONTEXT: INT (FK)
T1: VARCHAR
T2: VARCHAR
T3: VARCHAR
T4: VARCHAR
T5: VARCHAR
T6: VARCHAR
T7: VARCHAR
T8: VARCHAR
T9: VARCHAR
T10: VARCHAR
T1_result: TEXT
T2_result: TEXT
T3_result: TEXT
T4_result: TEXT
T5_result: TEXT
T6_result: TEXT
T7_result: TEXT
T8_result: TEXT
T9_result: TEXT
T10_result: TEXT
IS_SYNT_HEAD: VARCHAR
NATURE_TESTS: VARCHAR
Table_TESTS_FKIndex1
Table10_CONTEXT_ID_CONTEXT
Table10_CONTEXT
ID_CONTEXT: INT
Table5_UNITES_LEXICALES_NOMS_ID_UL_NOMS: INT (FK)
TAG: VARCHAR
CONTEXT: TEXT
NIV_ANNOTATION: INT
Table_CONTEXT_FKIndex1
Table5_UNITES_LEXICALES_NOMS_ID_UL_NOMS
Rel_context_tests (1/1)

Bibliographie

Bisetto, A. et Melloni, C. (2008). On the Interpration of Nominals : Towards e Result-Oriented Verb Classification. Publié dans *Proceedings of the 40th Linguistics Colloquium*, Frankfurt.

Fradin, B. (à paraître). Les nominalisations et la lecture 'moyen'. *Lexique*.

Fradin, B. et Kerleroux, F. (2003). Quelles bases pour les procédés de la morphologie constructionnelle ? Publié dans Fradin, B., Dal, G., Hathout, N., Kerleroux, F., Plénat, M., et Roché, M., éditeurs, *Les unités morphologiques, Silexicales, Vol.3*. CNRS & Université Lille 3, Villeneuve d'Ascq.

Godard, D. et Jayez, J. (1996). Types nominaux et anaphores : le cas des objets et des événements. Publié dans Mulder, W. D., Ryck, L. T.-D., et Vetters, C., éditeurs, *Cahiers Chronos 1*.

Polguère, A. (2003). *Lexicologie et sémantique lexicale*. Presses de l'université de Montréal, Montréal.

Pustejovsky, J. (1995). *The Generative Lexicon*. MIT Press, Cambridge.

van den Eynde, K. et Mertens, P. (2003). La valence : l'approche pronominale et son application au lexique verbal. *Journal of French Language Studies*, 13 :63–104.