



HAL
open science

Termes complexes et langues de spécialité en sciences humaines et sociales : que nous apprennent les textes intégraux ?

Evelyne Jacquey, Laurence Kister, Mario Marcon, Sabine Barreaux

► To cite this version:

Evelyne Jacquey, Laurence Kister, Mario Marcon, Sabine Barreaux. Termes complexes et langues de spécialité en sciences humaines et sociales : que nous apprennent les textes intégraux ?. Meta : journal des traducteurs, 2018, 63 (1), pp.7-29. 10.7202/1050512ar . hal-01875377

HAL Id: hal-01875377

<https://hal.science/hal-01875377>

Submitted on 17 Sep 2018

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Termes complexes et langues de spécialité en sciences humaines et sociales : que nous apprennent les textes intégraux ?

EVELYNE JACQUEY

Atilf UMR 7118 CNRS/Université de Lorraine, Nancy, France

Evelyne.Jacquey@atilf.fr

LAURENCE KISTER

Atilf UMR7118 Université de Lorraine, Nancy, France

Laurence.Kister@univ-lorraine.fr

MARIO MARCON

Atilf UMR 7118 CNRS/Université de Lorraine, Nancy, France

Mario.Marcon@atilf.fr

SABINE BARREAUX

Inist CNRS, Vandœuvre-lès-Nancy, France

Sabine.Barreaux@inist.fr

RÉSUMÉ

A partir d'un corpus d'articles scientifiques de trois disciplines de sciences humaines (archéologie, linguistique et psychologie), l'étude que nous présentons a pour objectif d'examiner la composition des candidats termes poly-lexicaux de forme *N_Adj* (*langage humain, niveau individuel, origine sociale, structure syllabique*). Nous nous concentrons sur les occurrences de candidats termes considérées comme relevant d'un usage disciplinaire par les annotateurs experts de chaque discipline. Nous nous intéressons, en particulier, aux classements possibles des noms et des adjectifs dans trois lexiques en interaction dans l'écrit scientifique : le lexique scientifique transdisciplinaire, les ressources terminologiques et le lexique de la langue générale. Les candidats termes analysés ont au moins un composant, le nom ou l'adjectif, qui appartient à au moins deux des trois disciplines du corpus. A l'issue de cette étude, nous constatons le rôle primordial du lexique disciplinaire et ses interactions fréquentes avec le lexique scientifique transdisciplinaire dans la formation des termes poly-lexicaux de forme *N_Adj*.

ABSTRACT

Our study is based on a specialized corpus of scientific papers dealing with three scientific fields of Humanities: Archeology, Linguistics and Psychology. The study aims at examining how transdisciplinary scientific lexicon and the disciplinary lexicon of each scientific field can produce new noun_adjective term candidates, for instance, *human language, individual level, social background, syllabic structure*. We focus on all of the occurrences of noun_adjective term candidates that expert annotators considered as terms in each field. Moreover, the selected candidates contain at least one component, the noun or the adjective, occurring in at least two fields. The results show the key role of the disciplinary lexicon and its recurrent interactions with the transdisciplinary scientific lexicon for creating noun_adjective terms.

RESUMEN

A partir de un corpus de artículos científicos de tres disciplinas de ciencias humanas (arqueología, lingüística y psicología), el objetivo del estudio que presentamos es examinar la composición de los candidatos a términos poliléxicos de forma *N_Adj* (*lenguaje humano, nivel individual, origen social, estructura silábica*). Nos enfocaremos en las incidencias de candidatos a términos que los anotadores expertos de cada disciplina consideran propias de un uso disciplinario. Nos centraremos con mayor énfasis en las clasificaciones posibles de los nombres y de los adjetivos en tres léxicos que interactúan entre sí en el escrito científico: el léxico científico interdisciplinario, los recursos terminológicos y el léxico de la lengua general. Los candidatos a términos analizados tienen por lo menos un componente, el nombre o el adjetivo, que pertenecen por lo menos a dos de cada tres disciplinas del corpus. Al final de este estudio, comprobamos el papel primordial del léxico disciplinario y sus interacciones frecuentes con el léxico científico interdisciplinario para la formación de los términos poliléxicos de forma *N_Adj*.

MOTS-CLÉS/KEYWORDS/PALABRAS CLAVE

lexique scientifique transdisciplinaire, terminologie, écrit scientifique, candidat terme *N_Adj*
scientific transdisciplinary lexicon, terminology, scientific discourse, term candidate, pattern *N_Adj*
léxico científico interdisciplinario, terminología, escrito científico, candidato a término *N_Adj*

Introduction

Notre objectif est d'étudier la composition des candidats termes poly-lexicaux de forme *N_Adj* en contexte tels qu'ils ont été relevés dans le cadre du projet TermITH (ANR-12-CORD-029 – Contint (Contenus numériques et interactions)) qui a réuni six partenaires : ATILF (coordination) LINA, LIDILEM, INIST, INRIA Grand-Est et INRIA Saclay, <www.atilf.fr/ressources/termith>. Plus précisément, nous cherchons à mettre en évidence des caractéristiques particulières du nom et de l'adjectif qui favorisent la probabilité pour le candidat d'accéder au statut de terme.

Le lexique, et plus particulièrement les mots et les groupes de mots, a occupé une place importante dans les études terminologiques depuis l'essor de la Théorie générale de la terminologie (Wüster 1981), et ce, malgré l'orientation onomasiologique de cette approche.

Pour ce qui est des mots étudiés, le nom représente la partie du discours censée accéder au statut de terme car il permet de désigner un concept (Lerat 2009). La présence massive des noms dans les dictionnaires terminologiques est aussi observée par l'Homme (2004a). Toutefois, elle envisage des études terminologiques lexico-sémantiques portant sur les adjectifs, les verbes et les adverbes, qu'ils soient ou non en relation avec un nom. Cabré (1999) affirme que la quantité de termes nominaux dépasse la quantité de termes adjectivaux et de termes verbaux en langue de spécialité.

Pour ce qui est des groupes de mots, la forme *N_Adj* a retenu l'attention des chercheurs en raison de sa fréquence. Cabré (1999) souligne la productivité de certains patrons syntaxiques dont la forme *N_Adj* pour la création de poly-lexicaux. Selon l'Homme (2004a), la forme *N_Adj* est reconnue comme l'une des plus productives en français et comme un indice non-négligeable pour l'extraction terminologique sur corpus. En plus de la fréquence remarquable des termes poly-lexicaux de forme *N_Adj*, le rôle que jouent les adjectifs constitue une problématique importante pour analyser la structure interne des termes complexes (L'Homme 2004a et Mendès

2003). Daille (2001), Maniez (2002), Monceaux (1997), entres autres, se sont focalisés sur l'analyse du fonctionnement de l'adjectif de relation dans la construction des poly-lexicaux. L'Homme (2004b) s'intéresse aux adjectifs déverbaux et aux adjectifs dénominaux qu'elle regroupe dans la classe des adjectifs dérivés sémantiques (ADS) : la structuration terminologique est influencée par les liens que ces adjectifs entretiennent avec les bases nominales et verbales à partir desquelles ils sont construits. L'Homme et Jia (2015) postulent la nécessité d'inclure des combinaisons lexicales spécialisées à bases nominales formées d'un nom et d'un adjectif dans les dictionnaires, notamment pour les collocationnels qui incluent des adjectifs qualificatifs (caractéristiques liées à l'utilisation : *logiciel convivial*), des adjectifs dénominaux (caractéristiques liées à l'interaction : *éditeur intuitif*) ou des adjectifs déverbaux (état particulier : *site évolutif*) et qui complètent un terme relevant d'un domaine de spécialité.

De plus, les campagnes d'annotation manuelle que nous avons menées au cours du projet TermITH ont mis au jour un nouvel aspect justifiant l'étude de ces termes : leurs composants peuvent recevoir différentes interprétations sémantiques qui relèvent de plusieurs lexiques présents dans les écrits scientifiques. Il s'agit des lexiques disciplinaires (*Ldisci*) propres à chaque domaine de spécialité, du lexique de la langue générale (*LG*) et du lexique scientifique transdisciplinaire (*LST*) commun à plusieurs disciplines scientifiques des sciences humaines et sociales.

L'attention que nous portons au lexique disciplinaire se fonde sur les aspects évoqués ci-dessus : le nom mais aussi l'adjectif qui composent le *N_Adj* peuvent appartenir au domaine de spécialité.

Le lexique de la langue générale nous intéresse dans la mesure où le processus de terminologisation peut conduire à une superposition ou une juxtaposition d'un sens spécialisé dans un domaine de spécialité et d'un sens commun en LG (Kocourek 1991). Ce phénomène peut être illustré par la polysémie du nom *cause* discutée par L'Homme (2004a). Ce nom signifie *ce qui produit un effet* en langue générale et *affaire pour laquelle une action est intentée en justice et qui fait l'objet d'un procès* dans le domaine du droit (les définitions sont extraites du TLFi, < atilf.atilf.fr >).

Le lexique transdisciplinaire retient notre attention car il joue un rôle primordial dans l'écrit scientifique en permettant la description et la présentation des activités scientifiques : il contient des noms comme *position, pratique, représentation* et des adjectifs comme *collectif, direct, général, humain, moyen, propre*. Ces unités lexicales sont essentielles à l'expression de la pensée scientifique (Tutin 2007 ; Drouin 2010a et b ; Hatier 2016 ; Hatier *et al.* 2014, Tutin et Grosmann 2014). Le lexique que nous utilisons pour nos travaux a été constitué à partir d'un corpus multidisciplinaire de 5 millions de mots et est composé de 500 textes relevant de dix disciplines de sciences humaines et sociales (50 textes par discipline) : anthropologie, économie géographique, histoire, linguistique, psychologie, sciences de l'éducation, sciences politiques, sciences de l'information et de la communication, sociologie. Il a été entièrement vérifié par le LIDILEM : Tutin (2007) est à l'origine de la première version de ce lexique en collaboration avec Drouin (2010a et b) de l'OLST. La version que nous utilisons est celle qui a été mise au point par Hatier (2013 et 2016) dans le cadre de sa thèse de doctorat.

Les candidats de forme *N_Adj* que nous avons étudiés sont, par exemple, *langage humain, facteurs personnels, décrochage scolaire, origine sociale, structure syllabique et orientation scolaire*. Les occurrences de candidats étudiés, dont quelques exemples sont retranscrits ci-dessous, sont extraites automatiquement à partir des articles scientifiques authentiques utilisés dans le cadre du projet TermITH et donc du corpus utilisé dans les travaux

présentés dans cet article. Dans ces exemples, les occurrences en italiques sont les seules proposées par la chaîne de traitement appliquée aux articles. Tout autre poly-lexical qu'un lecteur attentif pourrait envisager comme un candidat terme potentiel n'est pas en italique car il n'a pas été proposé par la chaîne de traitement.

Des travaux récents en phonétique-phonologie et en acquisition du langage montrent que l'étude de la syllabe est devenue un cadre nécessaire et incontournable pour comprendre le fonctionnement du langage et apporter des éléments intéressants de discussion à des questions et des problèmes d'ordre général non encore résolus : la syllabe est-elle l'unité fondamentale du *langage humain* ? Quels éléments pertinents pourraient appuyer l'existence phonétique, phonologique et psychologique de l'unité syllabique ? (Blaya et Fortin 2011)

Cette recherche permet de confirmer, dans la lignée des travaux précédent (Fortin et Al. 2004), que c'est la combinaison et le cumul d'un nombre important de *facteurs personnels*, familiaux et scolaires qui favorise le *décrochage scolaire*, qu'il n'y a pas une cause unique aux difficultés rencontrées et que cette combinaison varie entre les individus et les contextes. (Blaya et Fortin 2011)

D'autres recherches ont montré l'influence directe des stéréotypes sur le devenir des écoliers. Par exemple, Channouf, Magnard, Baudry et Perney (2005) se sont intéressés aux conséquences des stéréotypes liés à l'*origine sociale* sur les décisions d'*orientation scolaire*. (Desombre, Delelis, Lachal, Urban, Roye, Gaillet et Antoine 2008)

La fréquence d'une *structure syllabique* est donc liée au degré de complexité de l'attaque et/ou de la coda, celui-ci correspond au nombre de segments occupant cette position. (Vallée, Rousset et Bøe 2001)

Les trois lexiques que nous avons retenus (*disci*, *LST* et *LG*) permettent d'envisager neuf patrons ou combinaisons possibles pour les poly-lexicaux de forme *N_Adj*. Nous illustrons ces patrons par des exemples de candidats termes extraits de notre corpus de travail :

- *N_disci+Adj_disci* : *compétence sociale*
- *N_disci+Adj_LG* : *sépulture primaire*
- *N_disci+Adj_LST* : *restes humains*
- *N_LG+Adj_disci* : *mondes possibles*
- *N_LG+Adj_LST* : *face interne*
- *N_LG+Adj_LG* : *non représenté dans les corpus utilisés*
- *N_LST+Adj_disci* : *analyse textuelle*
- *N_LST+Adj_LG* : *étude détaillée*
- *N_LST+Adj_LST* : *différences significatives*

Une toute première observation d'un échantillon de notre corpus d'étude montrait une présence importante du *LST* dans les candidats termes, que ces derniers soient déjà présents ou non dans les ressources terminologiques. Lors de travaux antérieurs, Kister et Jacquy (2012) ont exploré les interactions entre le *LST* et les occurrences de termes et ont constaté qu'un élément du *LST* peut introduire un terme dans le cadre d'une dépendance directe lorsque le *N* est complément d'un *N* du *LST* comme dans *concept_{LST}* de la *dyslexie_{disci}* ou comme dans *la dyslexie est un trouble spécifique de l'acquisition_{LST} de la lecture_{disci}*. De plus, Jacquy et alii (2013) et Kister et alii (2015) ont constaté qu'en plus de la délimitation lexico-syntaxique, un élément du *LST* peut être intégré à un terme, notamment quand cet élément est un nom comme dans *évolution paléo-environnementale* (domaine de l'archéologie) ou *production langagière* (domaine de la linguistique).

Aussi, nous souhaitons approfondir le processus de terminologisation, encore peu exploré, pour lequel un élément du *LST* entre dans la formation d'un terme poly-lexical. Cela

revient à faire l'hypothèse que les *N_Adj* qui relèvent de la discipline scientifique du texte dont ils sont issus sont majoritairement construits avec un *N* relevant du *LST*. Parmi les trois patrons possibles en fonction de l'appartenance lexicale de l'adjectif, nous supposons que ce sont les adjectifs disciplinaires qui sont majoritaires.

Il s'ensuit que la vérification de cette hypothèse ouvrirait la voie à une réflexion sur le processus de terminologisation et sur la nature du *LST*. De plus, elle aurait un impact sur l'extraction terminologique automatique, notamment sur l'établissement et l'exploitation des ressources terminologiques mises à disposition des extracteurs.

Outre le rôle du *LST*, notamment en position nominale, nous avons pu vérifier l'influence majeure d'un composant disciplinaire lorsque celui-ci est un adjectif modifiant le nom. Dans ce cas, la présence de l'adjectif disciplinaire favorise l'accès du poly-lexical dans lequel il apparaît au statut de terme. La question est de savoir si l'influence d'un adjectif disciplinaire est la même quelle que soit l'appartenance lexicale du nom modifié. Nous nous intéressons en particulier à la productivité du patron *N_LG+Adj_disci* en guise de contre-vérification de l'acquis sur la terminologisation de *mots* de la *LG* et de la validité de notre hypothèse de contamination adjectivale (Kister *et alii*, 2015).

Après une présentation des données utilisées et de la méthodologie mise en œuvre, nous exposerons les résultats obtenus en termes de productivité des différents patrons de *N_Adj*. L'étude que nous présentons permet de constater que l'interaction entre *LST* et *Ldisci* produit un effet de contamination et favorise l'accès des poly-lexicaux au statut de terme.

1. Données de l'étude

Les données que nous avons utilisées proviennent de trois disciplines différentes réunies dans un corpus de 70 articles de revues : archéologie (11 articles), linguistique (29 articles) et psychologie (30 articles). Le corpus ainsi constitué contient 575 068 occurrences correspondant à 46 523 mots-formes (sur l'ensemble du corpus d'étude et non discipline par discipline) et 22 900 formes lemmatisées. Le tableau ci-dessous présente le corpus de manière plus détaillée en se concentrant sur sa distribution en termes de noms et d'adjectifs.

TABLEAU 1

Taille des corpus pour les 3 disciplines et en fonction des *N*, des *Adj* et des *N_Adj*

	Archéologie		Linguistique		Psychologie	
	Effectifs	Fréquences pour 1000	Effectifs	Fréquences pour 1000	Effectifs	Fréquences pour 1000
Total occurrences	89 651	-	223 338	-	262 169	-
Total mots différents	9 867	110.17	21 528	96.39	15 128	57.70
Nombre d'occurrences de noms	24 126	269.38	51 543	230.78	64 139	244.65
Nombre de noms différents	4 900	54.71	10 964	49.09	6 939	26.47
Nombre d'occurrences d'adjectifs	7 061	78.84	17 001	76.12	18 660	71.18

Nombre d'adjectif différents	2 089	23.32	4 546	20.35	3 032	11.57
Nombre d'occurrences de candidats termes <i>N_Adj</i>	3 735	41.7	8 730	39.08	10 530	40.16
Nombre de candidats de formes <i>N_Adj</i>	2 542	28.38	6 212	27.81	5 800	22.12

Comme le montre le tableau 1, la taille des sous-corpus varie en fonction des disciplines. Les sous-corpus de la linguistique et de la psychologie sont de taille comparable alors que le sous-corpus de l'archéologie est de taille nettement inférieure. Ceci n'est cependant pas problématique puisque nous mesurons la productivité des différents patrons de candidats termes en fréquence relative, calculée par rapport à la taille du sous-corpus de chaque discipline. Les fréquences relatives des adjectifs et des noms montrent que l'archéologie comporte plus d'éléments de ces deux catégories que la linguistique et la psychologie. Si on s'intéresse aux fréquences relatives des poly-lexicaux différents de forme *N_Adj*, l'archéologie et la linguistique sont très proches l'une de l'autre (23,38 % pour la première et 27,8 % pour la seconde) et que l'écart de ces deux disciplines avec la psychologie (22,12 %) est moindre par rapport à celui des autres mesures (taille des sous-corpus, nombre de noms et d'adjectifs).

Les textes ont subi plusieurs traitements. Ils ont été enrichis en annotations morpho-syntaxiques à l'aide de TreeTagger (Schmid 1994) et en occurrences de candidats termes. Les candidats termes projetés proviennent d'une terminologie extraite par la plateforme d'extraction multilingue TermSuite (Rocheteau et Daille 2011; Cram et Daille 2016). La projection de la liste des candidats s'effectue par correspondance entre candidats termes et segments textuels sur la base de leur forme lemmatisée et étiquetée grammaticalement. Les occurrences de candidats termes peuvent être ambiguës car une lecture disciplinaire et une lecture non-disciplinaire restent possibles en raison du mode de reconnaissance appliqué. C'est le cas du candidat mono-lexical *sujet* qui est proposé par TermSuite pour la linguistique. Dans une phrase de la forme *le participe passé des verbes se construisant avec l'auxiliaire être doit s'accorder en genre et en nombre avec le sujet de la phrase*, le candidat *sujet* fait l'objet d'un emploi disciplinaire. En revanche, dans un contexte comme *le sujet de la communication s'inscrit dans [...]*, l'occurrence de *sujet* reçoit une interprétation non-disciplinaire. Pour lever ce type d'ambiguïté, une évaluation manuelle de l'interprétation effective de chaque occurrence de candidat terme a été réalisée par des linguistes de l'ATILF et des documentalistes de l'INIST experts dans les différentes disciplines du corpus. De cette manière, l'ensemble des occurrences de candidats termes sont filtrées manuellement pour sélectionner les occurrences qui relèvent d'un usage effectivement disciplinaire et celles qui n'en relèvent pas. En traitement automatique de la langue, cette procédure de filtrage s'apparenterait à une classification permettant de distinguer les occurrences disciplinaires et les occurrences non-disciplinaires.

Suite au traitement automatique des textes et au filtrage manuel des occurrences de candidats termes, les enrichissements des données sont stockés au format XML-TEI complété par la proposition StandOff-Proposal, en cours de validation par le consortium TEI. La StandOff-Proposal permet de dissocier aisément le contenu textuel et l'ensemble des annotations. Ces traitements sont réalisés à l'aide de la plateforme de traitement spécifique au projet TermITH qui a fait l'objet d'un dépôt sur GitHub <<https://github.com/simonmeoni/termITH-tool>>.

A partir des données enrichies, trois critères ont guidé l'extraction des noms et des adjectifs qui composent les *N_Adj* examinés :

- ils sont communs à au moins deux des trois disciplines prises en compte dans l'expérience ;
- ils apparaissent dans des candidats termes de forme *N_Adj* ;
- les occurrences de ces candidats termes ont été jugées disciplinaires lors du filtrage manuel.

Le tableau 2 donne une vue quantitative des données sur lesquelles l'étude a été menée. Il indique que le filtrage manuel a été mis en œuvre sur presque cinq fois plus d'occurrences de candidats termes de forme *N_Adj* en psychologie (5151) qu'en archéologie (1526) et en linguistique (1536). Si on s'intéresse aux occurrences des candidats termes jugées disciplinaires, on s'aperçoit qu'elles sont presque trois fois plus nombreuses en psychologie (3101) qu'en archéologie (1187) et en linguistique (1111).

TABLEAU 2

Répartition des occurrences de candidats termes *N_Adj* en emplois disciplinaires et non-disciplinaires

	Archéologie	Linguistique	Psychologie
Occurrences jugées disciplinaires	1 187	1 111	3 001
Occurrences jugées non disciplinaires	339	425	2 150
Total	1 526	1 536	5 151

Le tableau 3 compare, d'une part, les proportions d'occurrences de candidats termes de forme *N_Adj* en fonction du nombre total de séquences *N_Adj* pour chaque discipline, et d'autre part, la proportion d'occurrences jugées disciplinaires parmi celles-ci.

TABLEAU 3

Proportions des candidats termes de forme *N_Adj* disciplinaire par discipline

	Archéologie	Linguistique	Psychologie
Candidats termes <i>N_Adj</i> /séquences <i>N_Adj</i> par discipline	60.03%	24.73%	88.81%
Occurrences jugées disciplinaires	77.79%	72.33%	41.74%

Le tableau 3 montre une nette variation dans les proportions de candidats termes de formes *N_Adj* par rapport au total des candidats termes proposés par l'extracteur terminologique TermSuite. La proportion de candidats termes de forme *N_Adj* est plus importante en psychologie qu'en archéologie qui est elle-même plus importante qu'en linguistique. Pour détecter les candidats termes, TermSuite utilise des règles linguistiques et des métriques statistiques et, parmi elles, les fréquences relatives calculées dans un corpus de spécialité et dans un corpus de langue générale (corpus de presse, *Le Monde*, de 82 millions de tokens (Daille 2017)). On peut donc émettre l'hypothèse qu'en linguistique le nombre de candidats termes identifiés par TermSuite est moins important qu'en psychologie et en archéologie parce qu'il y a moins de poly-lexicaux de forme *N_Adj* dont la fréquence relative en linguistique est significativement différente de leur fréquence relative en langue générale. Ainsi, si on observe les poly-lexicaux en linguistique du point de vue de leur fréquence relative, on peut supposer que leur usage est moins spécifique que celui qu'on a en archéologie et en psychologie. Pour ce qui est de la proportion d'occurrences de candidats termes *N_Adj* jugées disciplinaires, l'ordre change : environ 78 % des occurrences sont jugées disciplinaires en archéologie, 72 % en linguistique et seulement un peu plus de 41 % en psychologie. Cela montre que les *N_Adj* en archéologie et en linguistique sont beaucoup plus précis qu'en psychologie. Les *N_Adj* en psychologie ne renvoient probablement pas à des concepts stables mais représentent des

cooccurrences ou des collocations propres au discours de la psychologie, cooccurrences et collocations qui ne sont pas considérées comme disciplinaires par les experts de la psychologie.

Nous avons ensuite procédé à l'extraction des contextes de toutes les occurrences de candidats termes jugées disciplinaires dans chacune des trois disciplines. Le résultat de cette extraction est stocké dans un tableur avec, pour chaque occurrence, le fichier dont l'occurrence a été extraite, les identifiants des *tokens* qui la composent afin de pouvoir la situer précisément dans le texte, l'annotation manuelle, son contexte gauche et son contexte droit à l'intérieur du paragraphe (balise <p/>) où elle apparaît. Le tableau 4 présente un exemple des informations extraites pour le candidat *allomorphie radicale* en linguistique.

TABLEAU 4

Exemple du candidat *allomorphie radicale* en linguistique

ID Fichier	Identifiants occurrences	Catégories	Filtrage manuel	Contexte gauche	Candidat	Contexte droit
Linguistique_12-0089112_tel8425	#15676 #15667	NOM ADJ	Disciplinaire	Les	allomorphies radicales	sont la manifestation la plus visible de l'irrégularité massive et multiforme de la morphologie du français

2. Méthodologie

Pour l'ensemble des occurrences sélectionnées, nous avons ensuite vérifié l'annotation manuelle dans les rares cas où le filtrage manuel ne semblait pas cohérent. Nous avons rencontré de tels cas en archéologie et en psychologie. Pour l'archéologie, cela provient de confusions manifestes avec des usages transdisciplinaires comme pour *données nouvelles*.

Malgré une vision très partielle, ce site [moyenne vallée de l'Orb] fournit des *données nouvelles* sur les modes d'occupation des sols et sur les faciès céramique de la transition entre l'âge du Bronze et l'Âge du Fer. (Maziere 2001)

Pour la psychologie, les incohérences rencontrées sont dues au recouvrement avec d'autres disciplines comme, par exemple, avec les sciences de l'éducation. C'est le cas du candidat terme *sections scientifiques*. Ce candidat terme appartient à la terminologie produite par TermSuite, au même titre que *milieu scolaire*, par exemple. Cependant, *milieu scolaire* semble faire partie du lexique de la psychologie alors que *section scientifique* relève d'un usage beaucoup plus ponctuel qu'on retrouve essentiellement dans les plaquettes d'orientation, par exemple, et qui est plus utilisé par enseignants des classes de 3^{ème} ou de terminale que par les psychologues. On peut aussi noter que le candidat terme *section scientifique* ne fait pas partie du vocabulaire de référence utilisé en psychologie.

Au cours de deux enquêtes, la première conduite en 1994 et la seconde en 2006, nous avons évalué les sentiments d'efficacité personnelle (SEP) relatifs aux différents types d'études et de professions de lycéennes et de lycéens de terminales de *sections scientifiques* (S) et économiques et sociales (ES). Ces SEP varient à la fois en fonction du sexe et de la section de terminale fréquentée. (Blanchard, Lallemand et Steinbruckner 2009)

Ensuite, pour déterminer la contribution du lexique scientifique transdisciplinaire et du lexique disciplinaire dans la formation des termes de la forme *N_Adj*, nous avons confronté les noms et les adjectifs sélectionnés avec les ressources de référence du projet TermITH : le lexique scientifique transdisciplinaire et les ressources terminologiques de chacune des trois disciplines.

Les ressources terminologiques de référence pour chaque discipline ont été constituées par l'INIST. Elles proviennent du vocabulaire utilisé pour l'indexation des bases de données bibliographiques Francis et Pascal.

Pour opérer cette confrontation, les noms et les adjectifs sélectionnés sont recherchés par leur forme dans le lexique transdisciplinaire et dans les ressources terminologiques. Lorsqu'ils sont absents de ces lexiques, ils sont considérés comme relevant de la langue générale et reçoivent l'étiquette *LG*. Lorsqu'ils apparaissent dans le lexique scientifique transdisciplinaire ou dans les ressources terminologiques, leurs sens, tels qu'ils sont décrits explicitement dans le lexique scientifique transdisciplinaire ou tels qu'ils peuvent être inférés dans les ressources terminologiques, sont comparés manuellement au sens qu'ils ont dans l'occurrence *N_Adj* candidate où ils figurent. Lorsque seul le sens transdisciplinaire est jugé compatible, l'occurrence du nom ou de l'adjectif examiné est comptée comme transdisciplinaire. Le candidat *taille moyenne* est un exemple de ce type en archéologie. Le nom *taille* appartient au lexique scientifique transdisciplinaire avec le sens *grandeur* et apparaît aussi dans la ressource terminologique de l'archéologie comme élément du terme *atelier de taille*. Dans le candidat *taille moyenne*, c'est clairement le sens transdisciplinaire qui est sélectionné.

Equus mosbachensis palustris, Bonifay, 1980 : défini à Lunel-Viel (Hérault), dans les ensembles supérieur et inférieur, cet Equidé est représenté par un nombre important de restes. La présence simultanée d'espèces archaïques et évoluées au sein des niveaux archéologiques suggère, pour le gisement, un âge proche de la fin du Pléistocène moyen autour de 350ka (Fosse 19994). Cet Equidé a pour caractéristique un crâne de *taille moyenne* avec une gouttière nasale visible, ce qui est un caractère ancien. (Langlois 2005)

La même décision est prise du côté disciplinaire lorsque seul le sens terminologique est compatible avec celui de l'occurrence examinée. En linguistique, le candidat *objet direct* est un exemple parlant. Le nom *objet* apparaît dans le lexique scientifique transdisciplinaire avec le sens *objet scientifique* et est présent dans la ressource terminologique de la linguistique. Dans ce cas, nous avons opté pour le sens terminologique en raison de la combinaison avec l'adjectif *direct* qui conduit naturellement à la désignation d'un concept linguistique.

Mais il faut d'abord noter que ces propriétés sont largement indépendantes de la nature des liens sémantiques ou pragmatiques qui unissent les positions en jeu : il peut exister une transformation reliant V à T comme il peut en exister une entre la position d'*objet direct* et la position initiale dans une interrogative (Quel livre lis-tu (quel livre) ?) ; leur justification fonctionnelle est clairement différente dans un cas et dans l'autre ; les contraintes formelles qui pèsent sur elles sont identiques. (Bellier 2003)

Pour le candidat *classes moyennes* en psychologie, le sens transdisciplinaire du nom *classe* avec le sens *groupe – partie de*, de même que le sens disciplinaire tel qu'il apparaît dans le terme *classe sociale* dans les ressources terminologiques de la psychologie, sont compatibles avec le sens sélectionné par le candidat *classes moyennes*. Nous avons, dans ce cas, décidé de suivre le choix fait par l'expert lors du filtrage manuel : le sens disciplinaire.

Deux interrogations majeures peuvent être formulées à l'encontre de ces études consacrées au processus de socialisation de genre. D'une part, elles ont peu examiné les différences de représentations et comportements liés au genre selon les milieux socioculturels, la plupart des recherches portant sur des *classes moyennes*. (Mieyaa, Rouyer et Le Blanc 2012)

Mais lorsqu'il s'agit d'une erreur manifeste de jugement, nous avons opté pour le sens transdisciplinaire comme, par exemple, pour *manière* dans *manière directe* en psychologie. On

peut noter que le cas de *manière directe* en psychologie est très similaire à celui de *données nouvelles* en archéologie, cité ci-dessus.

Concernant la première question, les études sur les processus de traitement des mots écrits ont montré que les adultes illettrés diffèrent qualitativement des enfants de même niveau de lecture. En lecture, ils ont plus de difficultés pour utiliser la voie phonologique de décodage des mots que pour lire les mots de *manière directe* et globale. (Eme, Nantes et Delliaux 2011)

Une vue synthétique des proportions des catégories obtenues pour les noms et les adjectifs à l'issue de cette seconde lecture est fournie dans le tableau 5.

TABLEAU 5

Répartition des noms et des adjectifs dans les trois catégories de lexiques par discipline.

	3 disciplines		Archéologie		Linguistique		Psychologie	
<i>N_LST</i>	16	100%	33	91.67%	47	88.68%	49	84.48%
<i>N_disci</i>	0	0%	2	5.56%	5	9.43%	4	6.9%
<i>N_LG</i>	0	0%	1	2.77%	1	1.89%	5	8.62%
Total N	16		36		53		58	
<i>Adj_LST</i>	4	80%	14	63.64%	13	81.25%	14	60.87%
<i>Adj_disci</i>	1	20%	6	2.27%	3	18.75%	9	39.13%
<i>Adj_LG</i>	0	0%	2	9.09%	0	0%	0	0%
Total Adj	5		22		16		23	
<i>N+Adj_LST</i>	20	95.24%	47	81.03%	60	86.96%	63	77.78%
<i>N+Adj_disci</i>	1	4.76%	8	13.79%	8	11.59%	13	16.05%
<i>N+Adj_LG</i>	0	0%	3	4.91%	2	1.45%	5	6.17%
Total <i>N_Adj</i>	21		58		70		81	

On constate une présence notoire du lexique scientifique transdisciplinaire pour les noms comme pour les adjectifs. Ce lexique est présent pour plus de 95 % des noms et des adjectifs partagés par les trois disciplines. L'appartenance d'un nom ou d'un adjectif à l'une des trois disciplines se fonde sur la discipline dont relève le texte dans lequel figure ce nom ou cet adjectif. La discipline d'un texte a été déterminée lors de la constitution du corpus par des experts de chacune des disciplines prises en compte dans le projet TermITH. Parmi les noms figurant dans des candidats termes *N_Adj* partagés par les trois disciplines examinées, on trouve, par exemple :

- *analyse* comme dans *analyse fonctionnelle* (archéologie), *analyse conversationnelle* (linguistique), *analyses multivariées* (psychologie) ;
- *étude* comme dans *études biométriques* (archéologie), *étude linguistique* (linguistique), *études universitaires* (psychologie) ;
- *forme* comme dans *forme ovoïde* (archéologie), *forme nominale* (linguistique), *forme identitaire* (psychologie) ;
- *niveau* comme dans *niveau archéologique* (archéologie), *niveau textuel* (linguistique), *niveau individuel* (psychologie) ;
- *structure* comme dans *structure archéologique* (archéologie), *structure syllabique* (linguistique), *structure identitaire* (psychologie).

Pour les adjectifs présents dans des candidats termes *N_Adj* on peut remarquer les exemples suivants :

- *humain* comme dans *restes humains* (archéologie), *langage humain* (linguistique), *sujet humain* (psychologie) ;

Nombre d'occurrences du patron	27	2	17	10	-	3	129	22	19	229
<i>N_Adj disci</i>	27	2	17	10	-	3	126	16	3	204
<i>N_Adj non-disci</i>	0	0	0	0	-	0	3	6	16	25
Proportion de <i>N_Adj disci</i>	100%	100%	100%	100%	-	100%	97.67%	72.72%	15.79%	89.08%
Proportion de <i>N_Adj non-disci</i>	0%	0%	0%	0%	-	0%	2.33%	27.27%	84.21%	10.92%
Part patron/total occurrences	11.79%	0.86%	7.62%	4.37%	-	1.31%	56.33%	9.6%	8.29%	100%
Proportion de <i>N_Adj disci/N_Adj</i>	11.79%	0.86%	7.62%	4.37%	-	1.31%	55.6%	6.98%	1.31%	89.08%
Proportion de <i>N_Adj non-disci/N_Adj</i>	0%	0%	0%	0%	-	0%	1.29%	2.62%	6.98%	10.92%

En archéologie, comme pour l'ensemble du corpus, le patron mettant en jeu un nom et un adjectif avec tous deux un sens relevant de la langue générale est absent (*N_LG+Adj_LG*). Les huit autres patrons sont représentés même si leur représentation est parfois très faible. Ainsi, nous notons seulement 3 occurrences de *N_Adj* comportant un nom avec un sens de langue générale et un adjectif avec un sens transdisciplinaire (*N_LG+Adj_LST*) : 1,31 % des *N_Adj* examinés en archéologie.

Le tableau 6 montre la position dominante du patron *N_LST+Adj_disci* (par exemple, *analyse tracéologique*) : 55,60 % des occurrences de *N_Adj* examinées en archéologie relèvent de ce patron et sont disciplinaires.

D'après leur position dans la chaîne opératoire et en tenant compte de paramètres méthodologiques et taphonomiques, le nombre d'éclats préférentiels extraits de leur lieu de débitage se monterait à 67, parmi lesquels on trouve 19 pièces effectivement utilisées d'après l'*analyse tracéologique*. (Vallin, Masson, Caspar et Depiereux 2006)

La productivité du patron *N_LST+Adj_disci* (55,60 % du total des occurrences de candidats termes *N_Adj* en archéologie) doit être appréhendée en fonction du maximum possible qui est de 88 % du total initial. Ce maximum de 88 % résulte de la seconde lecture que nous avons effectuée sur le du filtrage manuel, au cours duquel nous avons éliminé les occurrences étiquetées disciplinaires dont nous avons jugé l'évaluation en première lecture erronée. Les corrections effectuées ont conduit au rejet de 25 occurrences jugées disciplinaires lors du filtrage manuel initial, ce qui entraîne un total d'occurrences disciplinaires de 204 sur les 229 (soit 87,93 %) de la sélection initiale dans les données du projet.

Le second patron le plus représenté, ce qui n'est pas surprenant, est le patron *N_disci+Adj_disci* (par exemple, *perforations naturelles*) tandis que le patron *N_disci+Adj_LST* (par exemple, *datation relative*) occupe la troisième position.

Parmi les objets de parure de l'enfant, nous avons repéré un humérus de lagomorphe et une vertèbre de poisson portant des *perforations naturelles* (fig. 2G-h). Les traces d'ocre sur ces pièces et l'utilisation d'un autre os de lagomorphe comme objet de parure indiquent que ces deux éléments pourraient également faire partie de la parure de l'enfant. (Vanhaeren et D'Errico 2001)

Le genre Equus est donc de première importance pour la *datation relative* des sites ans lesquels il est présent. (Langlois 2005)

Notre hypothèse initiale évoquée dans l'introduction était que parmi les neuf patrons qu'il est possible de construire à l'aide de noms et d'adjectifs relevant des lexiques disciplinaire, scientifique transdisciplinaire et de langue générale, les poly-lexicaux de forme *N_Adj* qui comportent un *N* appartenant au *LST* sont majoritaires et parmi ces patrons ceux qui comportent un *Adj* relevant du lexique *disci* sont les plus nombreux. Ainsi les poly-lexicaux qui accèdent le plus aisément au statut *disciplinaire* sont les *N_LST+Adj_disci* suivi des *N_Adj* de forme *N_LST+Adj_LST* puis des *N_Adj* de forme *N_LST+Adj_LG*.

Les données analysées en archéologie confirment en partie cette hypothèse avec une place prépondérante occupée par le patron *N_LST+Adj_disci*. La conclusion intermédiaire à laquelle conduisent les données de l'archéologie serait la suivante : le *LST* joue un rôle déterminant dans la construction des termes de forme *N_Adj* lorsque la tête nominale admet une interprétation transdisciplinaire. Quand la tête nominale admet une interprétation disciplinaire, c'est l'appartenance de l'adjectif au lexique disciplinaire qui devient alors productive.

Dernier point sur l'archéologie, en complément des observations précédentes centrées sur les *N_Adj* disciplinaires les proportions de patrons majoritaires dans la production d'occurrences de candidats *N_Adj* non-disciplinaires ont été calculées sur les occurrences rejetées lors du processus de correction du filtrage manuel (seconde lecture). Ce calcul a été réalisé à titre indicatif à ce stade de l'analyse et sera complété par une analyse ultérieure du même type sur les occurrences jugées non-disciplinaires lors du filtrage manuel (première lecture). Pour les *N_Adj* non-disciplinaires, nous constatons la position dominante du patron *N_LST+Adj_LST* (par exemple, *différences significatives*). Ce patron est suivi du patron *N_LST+Adj_LG*.

Bien que les restes de chevaux paraissent appartenir à un même taxon, il était nécessaire, avant tout, regroupement, de montrer qu'ils ne présentaient pas de *différences significatives* entre les différentes couches. (Langlois 2005)

3.2. Productivité des patrons des candidats termes en linguistique

Dans le domaine de la linguistique, les résultats sont plus marqués. On note tout d'abord que le nombre d'occurrences disciplinaires rejetées (deuxième lecture) est très faible, ce qui se traduit par le fait qu'un peu plus de 99 % des occurrences jugées disciplinaires lors du filtrage manuel (première lecture) sont conservées en seconde lecture.

TABLEAU 7

Productivité des neufs patrons de candidats termes possibles de forme *N_Adj* en linguistique

	<i>N_disci+Adj_disci</i>	<i>N_disci+Adj_LG</i>	<i>N_disci+Adj_LST</i>	<i>N_LG+Adj_disci</i>	<i>N_LG +Adj_LG</i>	<i>N_LG +Adj_LST</i>	<i>N_LST+Adj_disci</i>	<i>N_LST +Adj_LG</i>	<i>N_LST +Adj_LST</i>	Tous patrons
Nombre d'occurrences du patron	86	-	39	2	-	-	304	4	2	437
<i>N_Adj disci</i>	85	-	39	2	-	-	303	4	0	433
<i>N_Adj non-disci</i>	1	-	0	0	-	-	1	0	2	4
Proportion de <i>N_Adj disci</i>	98.84%	-	100%	100%	-	-	99.67%	100%	0%	99.08%

Proportion de <i>N_Adj non-disci</i>	1.16%	-	0%	0%	-	-	0.33%	0%	100%	0.92%
Part patron/total occurrences	19.68%	-	8.92%	0.46%	-	-	69.57%	0.92%	0.46%	100%
Proportion de <i>N_Adj disci/N_Adj</i>	19.45%	-	8.92%	0.46%	-	-	69.34%	0.92%	0%	99.08%
Proportion de <i>N_Adj non-disci/N_Adj</i>	0.23%	-	0%	0%	-	-	0.23%	0%	0.46%	0.92%

L'ordre de représentation des patrons en fonction de leur productivité constatée dans les occurrences disciplinaires est le même que pour l'archéologie : *N_LST+Adj_disci*, *N_disci+Adj_disci*, *N_disci+Adj_LST*. Les candidats *analyse conversationnelle* (*N_LST+Adj_disci*), *langue parlée* (*N_disci+Adj_disci*) et *position initiale* (*N_disci+Adj_LST*) sont des exemples représentatifs des patrons cités ci-dessus.

Ces courants relèvent en France comme ailleurs, de différents domaines – de l'analyse de discours (Roulet *et alii.*1985, Coulhard 1992, van Dijk 1997), à la pragmatique des interactions verbales (Kerbrat-Orecchioni 1990, 1992, 1994) à la logique interlocutoire (Trognon 1999), à l'*analyse conversationnelle* (Sacks 1992, Sacks, Schegloff et Jefferson 1974). (Mondada et Traverso 2005)

Une banque de données de corpus de *langue parlée* en interaction : CLAPI. (Mondada et Traverso 2005)
Si on se penche sur la place de ces groupements consonantiques dans la syllabe, nous constatons que la plupart (7 %) occupent la *position initiale*, contre 33 % en finale de syllabe. (Vallée, Rousset et Böe 2001)

En plus du patron *N_LG+Adj_LG*, deux autres patrons sont totalement absents : *N_disci+Adj_LG* et *N_LG+Adj_LST*.

3.3. Productivité des patrons de candidats termes en psychologie

La psychologie présente des résultats similaires à ceux déjà observés du point de vue de la productivité des patrons majoritaires. En plus du patron *N_LG+Adj_LG*, le patron *N_disci+Adj_LG* est totalement absent. De plus, même si l'ordre des patrons majoritaires est le même dans la production d'occurrences *N_Adj* disciplinaires, les écarts sont beaucoup moins importants par rapport aux autres disciplines. Le patron *N_LST+Adj_disci* (par exemple, *contexte scolaire*), 48,38 % des occurrences de *N_Adj* en psychologie, perd environ 20 points par rapport à la linguistique (69,34%) et environ 7 points par rapport à l'archéologie (55,60 %) ;

L'estime de soi a été abondamment étudiée dans le *contexte scolaire*. En particulier, un lien a été mis en évidence entre « estime de soi » et « performance scolaire » (une haute estime de soi étant associée à la réussite scolaire et une basse étant associées à l'échec scolaire). (Dozot, Piret et Romainville 2009)

Le patron *N_disci+Adj_disci* (par exemple, *école primaire*), 28,60 % des occurrences de *N_Adj* en psychologie, gagne environ 10 points par rapport à la linguistique (19,45 %) et environ 18 par rapport à l'archéologie (11,79 %)

Dès lors, la difficulté scolaire à l'*école primaire* peut être appréhendée à partir des réponses institutionnelles ou des besoins éducatifs particuliers. (Desombre, Delelis, Lachal, Urban, Roye, Gaillet et Antoine 2008)

Cette perte et ce gain de productivité pour les patrons *N_LST+Adj_disci* et *N_disci+Adj_disci* conduisent nécessairement à un resserrement des écarts entre patrons.

TABLEAU 8

Productivité des neuf patrons de candidats termes possibles de forme *N_Adj* en psychologie

	<i>N_disci+Adj_disci</i>	<i>N_disci+Adj_LG</i>	<i>N_disci+Adj_LST</i>	<i>N_LG+Adj_disci</i>	<i>N_LG+Adj_LG</i>	<i>N_LG+Adj_LST</i>	<i>N_LST+Adj_disci</i>	<i>N_LST+Adj_LG</i>	<i>N_LST+Adj_LST</i>	Tous patrons
Nombre d'occurrences du patron	308	-	98	95	-	23	523	3	27	1077
<i>N_Adj disci</i>	308	-	97	94	-	23	521	1	21	1065
<i>N_Adj non-disci</i>	0	-	1	1	-	0	2	2	6	12
Proportion de <i>N_Adj disci</i>	100%	-	98.98%	98.95%	-	100%	99.62%	33.33%	77.78%	98.89%
Proportion de <i>N_Adj non-disci</i>	0%	-	1.02%	1.05%	-	0%	0.38%	66.67%	22.22%	1.11%
Part patron/total occurrences	28.6%	-	9.1%	8.82%	-	2.14%	48.56%	0.28%	2.51%	100%
Proportion de <i>N_Adj disci/N_Adj</i>	28.6%	-	9.01%	8.73%	-	2.14%	48.38%	0.09%	1.95%	98.89%
Proportion de <i>N_Adj non-disci/N_Adj</i>	0%	-	0.09%	0.09%	-	0%	0.19%	0.19%	0.56%	1.11%

Ce resserrement des écarts peut sembler surprenant du fait de la grande quantité d'occurrences jugées disciplinaires lors du processus de correction, près de 99 %. Dans le cas de la psychologie, on note une présence importante de candidats pouvant relever d'autres disciplines scientifiques, comme la linguistique (*langage écrit*), les sciences de l'éducation (*enseignement secondaire*), la didactique (*communication orale*) et la sociologie (*catégorie sociale*), tant dans les ressources terminologiques que dans l'expertise des annotateurs qui ont réalisé le filtrage manuel.

Des études de psychologie du développement ont porté sur les processus de traitement des mots écrits chez les adultes faibles lecteurs en référence aux modèles et aux méthodes décrivant le développement du *langage écrit* chez l'enfant (Ehri 1987, Frith 1985, Stanovich, Siegel et Gottardo 1997), et ont montré des différences qualitatives entre adultes illettrés et jeunes lecteurs. (Eme, Nantes et Delliaux 2011)

Ainsi, Caille et O'Prey (2005), dans une étude portant sur des jeunes en fin d'*enseignement secondaire* ont montré que l'estime de soi des étudiants influence la réussite scolaire ultérieure (au baccalauréat) et l'attrait pour les projets d'études les plus ambitieuses (Bac+5) alors que leur passé scolaire semble avoir peu d'influence sur leur estime de soi. (Dozot, Piret et Romainville 2009)

Qu'en est-il dans les activités de *communication orale* ? Car même si la lecture est un facteur critique de réussite sociale et professionnelle, les capacités à communiquer en tant que telles sont également importantes pour la recherche d'un emploi et les interactions sociales. (Eme, Nantes et Delliaux 2011)

Depuis, l'accès à l'emploi des diplômés scientifiques dans leur ensemble s'est fortement dégradé (« Génération 2001 ») même s'ils semblent conserver globalement des avantages nets en termes de salaires, de statut de l'emploi et de *catégorie sociale* par rapport aux autres disciplines de l'enseignement supérieur (Béwudé *et al.* 2006). (Béwudé, Fourcade, Giret et Moulet 2007)

Dans le cadre de notre étude, un certain nombre de noms et d'adjectifs montrent que les ressources terminologiques ne sont pas suffisamment spécialisées lorsqu'on compare des domaines potentiellement connexes. On peut constater une richesse due à la prise en compte des différents domaines avec lesquelles la psychologie au sens strict entretient des relations sur le plan scientifique (sciences du langage, sciences de l'éducation, sociologie) ou pratique (orientation scolaire, universitaire et professionnelle, acquisition du langage, pédagogie, échec

scolaire), tant au niveau du corpus en psychologie que des ressources terminologiques de référence qui lui sont associées. Il faut rappeler ici que le corpus TermITH ainsi que tous les traitements qui lui ont été appliqués avaient pour objectif d'identifier les créations terminologiques en cours de stabilisation, objectif secondaire du projet qui visait prioritairement l'indexation automatique des textes. Ainsi, la diversité qu'on constate dans le corpus de psychologie n'est pas surprenante : elle est conforme aux objectifs du projet pour lesquels le corpus a été constitué. La diversité constatée dans les ressources terminologiques de référence en psychologie, quant à elle, pose problème lorsqu'on cherche à mesurer le caractère disciplinaire ou non d'un candidat terme. Pour pallier cette difficulté, une piste possible serait de subdiviser les ressources en sous-domaines en tenant compte de la question des termes migrants (Toma 2002) et en limitant le chevauchement entre sous-domaines. Pour cela, il serait nécessaire de faire appel à des experts pour la modélisation de chaque sous-domaine.

3.4. Productivité des patrons des candidats termes *N_Adj* communs aux trois disciplines

Si on s'intéresse enfin aux occurrences de candidats termes *N_Adj* dont le nom ou l'adjectif est commun aux trois disciplines et appartient à un candidat terme, l'ordre des patrons prioritaires reste le même. La quantité d'occurrences disciplinaires conservées est tout à fait comparable à celles de la linguistique et de la psychologie. Pour les deux premiers patrons, *N_LST+Adj_disci* (par exemple, *caractéristique morphologique*) et *N_disci+Adj_disci* (par exemple, *relations sociales*), les écarts constatés se situent dans une position intermédiaire par rapport à ceux qui ont été observés pour la linguistique et la psychologie.

En ce qui concerne les chevaux de Lunel-Viel et de Caune de l'Arago et compte tenu des *caractéristiques morphologiques* et biométriques des éléments du squelette, nous pensons que ces chevaux font partie du groupe des grands chevaux de type mosbachensis. (Langlois 2005)

Ajoutons que Gergen définit aussi un « constructivisme social », celui de Vygotski ou de Bruner par exemple. Il affirme alors que « l'esprit » construit la réalité par sa relation au monde et *les relations sociales*. (Dumora et Boy 2008)

TABLEAU 9

Productivité des neuf patrons de candidats termes de forme *N_Adj* avec un *N* et/ou *Adj* commun(s) aux trois disciplines

	<i>N_disci+Adj_disci</i>	<i>N_disci+Adj_LG</i>	<i>N_disci+Adj_LST</i>	<i>N_LG+Adj_disci</i>	<i>N_LG +Adj_LG</i>	<i>N_LG +Adj_LST</i>	<i>N_LST+Adj_disci</i>	<i>N_LST +Adj_LG</i>	<i>N_LST +Adj_LST</i>	Tous patrons
Nombre d'occurrences du patron	164	-	35	18	-	-	374	14	7	612
<i>N_Adj_disci</i>	164	-	35	18	-	-	373	9	4	603
<i>N_Adj non-disci</i>	0	-	0	0	-	-	1	5	3	9
Proportion de <i>N_Adj_disci</i>	100%	-	100%	100%	-	-	99.7%	64.2%	57.1%	98.5%
Proportion de <i>N_Adj non-disci</i>	0%	-	0%	0%	-	-	0.27%	35.7%	42.8%	1.47%
Part patron/total occurrences	26.8%	-	5.72%	2.94%	-	-	61.1%	2.29%	1.14%	100%
Proportion de <i>N_Adj_disci/N_Adj</i>	26.8%	-	5.72%	2.94%	-	-	60.9%	1.47%	0.65%	98.5%

Proportion de N_{Adj} non-disci/ N_{Adj}	0%	-	0%	0%	-	-	0.16%	0.82%	0.49%	1.47%
--	----	---	----	----	---	---	-------	-------	-------	-------

3.5. Synthèse

Que les résultats soient examinés par discipline ou simultanément pour toutes les disciplines, les candidats termes examinés se répartissent selon huit des neuf patrons possibles (*cf.* Introduction) car le patron $N_{LG}+Adj_{LG}$ n'est pas représenté dans les corpus que nous utilisons. Par rapport à l'hypothèse de départ, nous constatons que deux des patrons censés produire des candidats N_{Adj} disciplinaires sont représentés et sont en position dominante : $N_{LST}+Adj_{disci}$ et $N_{disci}+Adj_{LST}$. Ces résultats mettent en avant le rôle primordial des interactions entre le lexique disciplinaire et le lexique transdisciplinaire ainsi que la contribution importante du lexique transdisciplinaire dans la construction des candidats termes disciplinaires ne figurant pas dans les ressources terminologiques. Les résultats obtenus montrent aussi l'autonomie du lexique disciplinaire par la seconde position en termes de productivité occupée par le patron $N_{disci}+Adj_{disci}$.

Il reste à noter que les résultats présentés jusqu'ici concernent les occurrences de candidats termes après élimination des termes qui figurent déjà dans les ressources terminologiques de référence. Nous avons comparé ces résultats avec ceux que nous aurions obtenus si nous n'avions pas procédé à cette élimination.

Dans les quatre tableaux ci-après (Tableaux 10 à 13), la partie gauche, intitulée *hors terminologie*, décrit les proportions de patrons de candidats termes après élimination des occurrences des candidats déjà présents dans les ressources terminologiques. Cette partie gauche est identique aux résultats présentés ci-avant (*cf.* 3.1 à 3.4). La partie droite, intitulée *avec terminologie*, montre les proportions obtenues sans procéder à cette élimination.

TABLEAU 10

Comparaison entre les productivités des patrons selon que l'on tient compte de la présence ou non des candidats termes dans l'une des trois terminologies

	3 disciplines hors terminologie			3 disciplines avec terminologie		
	Patrons/total occurrences	Patrons disci/total occurrences	Patrons non-disci/total occurrences	Patrons/total occurrences	Patrons disci/total occurrences	Patrons non-disci/total occurrences
$N_{disci}+Adj_{disci}$	31.08%	31.08%	0%	26.8%	26.8%	0%
$N_{disci}+Adj_{LST}$	5.33%	5.33%	0%	5.72%	5.72%	0%
$N_{LST}+Adj_{disci}$	58.39%	58.26%	0.13%	61.11%	60.95%	0.16%
$N_{LST}+Adj_{LG}$	1.82%	1.17%	0.65%	2.29%	1.47%	0.82%
$N_{LST}+Adj_{LST}$	0.91%	0.52%	0.39%	1.14%	0.65%	0.49%
Total	97.53%	96.36%	1.17%	97.06%	95.59%	1.47%

TABLEAU 11

Comparaison entre les productivités des patrons selon que l'on tient compte de la présence ou non des candidats termes dans la terminologie de l'archéologie

	archéologie hors terminologie			archéologie avec terminologie		
	Patrons/total occurrences	Patrons disci/total occurrences	Patrons non-disci/total occurrences	Patrons/total occurrences	Patrons disci/total occurrences	Patrons non-disci/total occurrences
$N_{disci}+Adj_{disci}$	11.24%	11.24%	0%	11.64%	11.64%	0%
$N_{disci}+Adj_{LST}$	7.75%	7.75%	0%	7.33%	7.33%	0%
$N_{LST}+Adj_{disci}$	57.36%	56.20%	1.16%	55.60%	54.31%	1.29%
$N_{LST}+Adj_{LG}$	9.69%	7.36%	2.33%	10.78%	6.9%	2.59%
$N_{LST}+Adj_{LST}$	8.14%	1.94%	6.2%	8.19%	1.29%	6.9%
Total	94.18%	84.49%	9.69%	93.54%	81.47%	10.78%

TABLEAU 12

Comparaison entre les productivités des patrons selon que l'on tient compte de la présence ou non des candidats termes dans la terminologie de la linguistique

	linguistique hors terminologie			linguistique avec terminologie		
	Patrons/total occurrences	Patrons disci/total occurrences	Patrons non-disci/total occurrences	Patrons/total occurrences	Patrons disci/total occurrences	Patrons non-disci/total occurrences
N_disci+Adj_disci	17.57%	17.38%	0.19%	19.68%	19.45%	0.23%
N_disci+Adj_LST	10.28%	10.28%	0%	8.92%	8.92%	0%
N_LST+Adj_disci	70.65%	70.47%	0.19%	69.57%	69.34%	0.23%
N_LST+Adj_LST	0.37%	0%	0.37%	0.46%	0%	0.46%
Total	98.87%	98.13%	0.75%	98.63%	97.71%	0.92%

TABLEAU 13

Comparaison entre les productivités des patrons selon que l'on tient compte de la présence ou non des candidats termes dans la terminologie de la psychologie

	psychologie hors terminologie			psychologie avec terminologie		
	Patrons/total occurrences	Patrons disci/total occurrences	Patrons non-disci/total occurrences	Patrons/total occurrences	Patrons disci/total occurrences	Patrons non-disci/total occurrences
N_disci+Adj_disci	36.87%	36.87%	0%	28.6%	28.6%	0%
N_disci+Adj_LST	7.44%	7.36%	0.08%	9.1%	9.01%	0.9%
N_LST+Adj_disci	44.46%	44.31%	0.15%	48.56%	48.38%	0.19%
N_LST+Adj_LG	0.23%	0.08%	0.15%	0.28%	0.09%	0.19%
N_LST+Adj_LST	2.05%	1.59%	0.46%	2.51%	1.95%	0.56%
Total	91.05%	90.21%	0.89%	89.05%	88.03%	1.84%

La comparaison montre que les proportions varient peu et que les tendances sont identiques. La distribution des patrons classés par proportions décroissantes reste stable que l'on prenne en compte ou non la présence des termes dans les ressources terminologiques. Ceci invite à supposer que les patrons de termes non encore référencés, et donc uniquement identifiables dans les textes intégraux, suivent des patrons similaires à ceux des candidats termes déjà référencés dans les ressources terminologiques. Nous aurons l'occasion de vérifier cette observation lorsque nous examinerons les résultats d'une évaluation manuelle ultérieure dont l'objectif est d'identifier les candidats termes de toutes formes, simple ou complexe, qu'il serait pertinent d'ajouter aux ressources terminologiques que nous avons utilisées pour ces travaux. L'évaluation manuelle sera réalisée par les experts qui ont participé à l'établissement des ressources terminologiques de référence de chaque discipline.

Conclusion

L'analyse du fonctionnement des candidats termes poly-lexicaux de forme *N_Adj* nous a permis de mesurer l'importance des différentes tendances lexicales qui existent dans l'écrit scientifique. Nous nous sommes en particulier focalisés sur le rôle joué par le lexique transdisciplinaire affirmé comme fondamental par de nombreux auteurs qui ont étudié le fonctionnement de l'écrit scientifique. Les différents patrons de candidats termes que nous avons repérés dans ce corpus montrent le rôle primordial du critère d'appartenance aux ressources terminologiques mais aussi la contribution du lexique transdisciplinaire dans la formation des termes de forme *N_Adj*. Pour l'ensemble des mesures effectuées (avec ou sans prise en compte de la préexistence de candidats termes dans les terminologies de domaines), nous constatons une forte représentation des *N_disci+Adj_disci* (deuxième rang) et une représentation dominante des *N_LST+Adj_disci*

(premier rang). La productivité de ces deux patrons s'oppose à la productivité beaucoup plus faible du patron *N_disci+Adj_LST*. Le rôle du lexique scientifique transdisciplinaire dans la construction des candidats termes poly-lexicaux est confirmé par la position dominante du patron *N_LST+Adj_disci*, mais l'étude menée montre que la contribution de ce sous-lexique de l'écrit scientifique est majeure lorsque le nom relève d'un usage transdisciplinaire dans le poly-lexical de forme *N_Adj*.

Cette première photographie de la répartition des *N_Adj* en fonction des critères d'appartenance du nom et de l'adjectif au lexique scientifique transdisciplinaire, aux ressources terminologiques et au lexique de la langue générale permet de commencer à repenser le rôle du lexique transdisciplinaire dans le processus de terminologisation et d'extraction automatique des termes. Les conclusions encore partielles de notre étude soulèvent trois pistes.

La première concerne l'analyse des occurrences des candidats termes *N_Adj* non-disciplinaires, jugées comme telles par les experts lors de la première phase d'enrichissement des données (première lecture) ou rejetés lors du processus de correction du filtrage manuel (seconde lecture). Cette analyse permettra de valider ou d'infirmer l'hypothèse d'une série de patrons privilégiés pour la construction de candidats termes, notamment ceux qui sont déjà présents dans les ressources terminologiques.

La deuxième porte sur la couverture d'une ressource terminologique. Celle-ci est toujours fonction de l'objectif de sa constitution. Dans le cas du projet TermITH, les buts sont l'indexation documentaire de domaines assez vastes et l'enrichissement de référentiels également vastes : la couverture doit donc être maximale, ce qui entraîne le croisement inévitable entre domaines. L'étude menée a ainsi montré que certains candidats termes semblent relever de domaines connexes, notamment dans le cas de la psychologie. Cela devrait amener à redéfinir la couverture éventuelle des ressources terminologiques pour décrire ces domaines et, par conséquent, à envisager une redéfinition des limites des domaines en SHS.

La troisième et dernière piste touche des candidats termes qui semblent relever du lexique scientifique transdisciplinaire et qui devraient entraîner un enrichissement de ce lexique, ce qui conduirait naturellement à un élagage des ressources terminologiques.

REFERENCES

- BELLIER Patrick (2003) : L'orientation générativiste. *LINX*, 48, 91-100.
- BEWUDE Catherine, FOURCADE Bernard, GIRET Jean-François et MOULLET Stéphanie (2007) : Etudes scientifiques et marché du travail : éléments de réflexion sur la crise des formations en sciences. *L'orientation scolaire et professionnelle*. 36(4), 503-532.
- BLANCHARD Serge, LALLEMAND Noëlle et STEINBRUCKNER Marie-Laure (2009) : L'évolution des sentiments d'efficacité scolaire et professionnels de lycéens français entre 1994 et 2006. *L'orientation scolaire et professionnelle*. 38(4), 417-449.
- BLAYA Catherine et FORTIN Laurier (2011) : Les élèves français et québécois à risques de décrochage scolaire : comparaison entre les facteurs de risques personnels, familiaux et scolaires. *L'orientation scolaire et professionnelle*. 40(1).
- CABRE Maria Teresa (1999) : *Terminology : theory, methods and applications*. Amsterdam/Philadelphia : John Benjamins.
- CRAM, Damien et DAILLE, Béatrice (2016) : Terminology Extraction with Term Variant Detection. In : *Proceedings of ACL-2016 System Demonstration*. (ACL-2016, Humbolt University, Berlin, 7-12 août 2016).
- DAILLE, Béatrice (2017) : *Term Variation in Specialized Corpora*. John Benjamins Publishing Compagny. DOI: 10.1075/tlrp.19, 272p.

- DAILLE, Béatrice (2001) : Qualitative Terminology Extraction : identifying relational Adjectives. In : Didier BOURIGAUT, Christian JACQUEMIN and Marie-Claude L'HOMME, dir. *Advances in computational Terminology*. Amsterdam : John Benjamins, 149-166.
- DESOMBRE Caroline, DELELIS Gérard, LACHAL Marc, URBAN Eugène, ROYE Louis, GAILLET Françoise et ANTOINE Laura (2008) : Stéréotypes de la difficulté scolaire : un outil de recueil. *L'orientation scolaire et professionnelle*.37(2).
- DROUIN, Patrick (2010a) : Extracting a bilingual transdisciplinary scientific Lexicon. In Magali PAQUOT and Sylviane GRANGER, dir. *Proceedings of eLexicography in the 21st century : new challenges, new applications*, vol.7, Louvain : Presses Universitaires de Louvain, 43-54.
- DROUIN, Patrick (2010b) : From a bilingual transdisciplinary scientific Lexicon to bilingual transdisciplinary scientific Collocations. In : Anne DYSKTRA and Tanneke SCHONHEIM, éd. *Proceedings on the 14th Euralex International Congress*. (14th Euralex International Congress, Fryske Akademy, Leeuwarden-Ljouwert, 6-10 July 2010).
- DOZOT Caroline, PIRET Anne et ROMAINVILLE Marc (2009) : L'estime de soi des étudiants de première année du supérieur en abandon d'études. *L'orientation scolaire et professionnelle*. 38(2), 205-230.
- DUMORA Bernadette et BOY Thierry (2008) : Les perspectives constructivistes et constructionnistes de l'identité (1^{ère} partie). *L'orientation scolaire et professionnelle*. 37(3) 347-36
- EME Elsa, NANTES Nicolas et DELLIAUX Christine. (2011) : Analyse cognitive et linguistique de l'illettrisme : bilan des études et implications pour la formation. *L'Orientation scolaire et professionnelle*. 40(3).
- HATIER Sylvain (2016) : *Identification et analyse linguistique du lexique scientifique transdisciplinaire. Approche ouillée sur corpus d'article de recherche en SHS*. Thèse de doctorat en sciences du langage, LIDILEM, Grenoble : Université Grenoble Alpes.
- HATIER, Sylvain (2013) : Extraction des mots simples du lexique scientifique transdisciplinaire dans les écrits de sciences humaines : une première expérimentation. In : *Actes de la 15^{ème} rencontre des étudiants chercheurs en informatique pour le traitement automatique des langues*. (Recital'2013, Les Sables d'Olonne, 17-21 juin 2013, 138-149).
- HATIER, Sylvain, TUTIN, Agnès, JACQUES, Marie-Paule, JACQUEY, Evelyne et KISTER, Laurence (2014) : Catégorisation sémantique des noms simples du lexique transdisciplinaire. In : *Etude de lexiques à vocation particulière : approches théoriques, méthodologiques, pédagogiques et multidisciplinaires*. (Congrès ACFAS, colloque 330, Université de Montréal, Montréal, 12-16 mai 2014).
- JACQUEY Evelyne, TUTIN Agnès, KISTER Laurence, JACQUES Marie-Paule, HATIER Sylvain et OLLINGER Sandrine (2013) : Filtrage terminologique parle lexique transdisciplinaire scientifique : une expérimentation en sciences humaines. In *Actes de la conférence TIA* (TIA 2013, Université de Paris 13, Villetaneuse, 28-30 octobre 2013).
- KISTER, Laurence et JACQUEY Evelyne (2012) : Relations syntaxiques entre lexique terminologique et transdisciplinaire : analyse en texte intégral. In : *Actes de la conférence CMLF* (CMLF 2012, Université de Lyon, Lyon, 4-7 juillet 2012).
- KISTER, Laurence, JACQUEY, Evelyne, MARCON, Mario et BARREAUX Sabine (2015) : Indices lexico-syntaxiques pour la reconnaissance des termes de la forme *N_Adj* dans les textes intégraux. In : *Actes TOTh 2015*. (TOTh 2015, Université de Savoie, Chambéry, 4-5 juin 2015).
- KOKOUREC, Rostislav (1991) : *La langue française de la technique et de la science. Vers une linguistique de la langue vivante*. Wiesbaden : Oscar Brandstetter.
- LANGLOIS Anne (2005) : Le Cheval du gisement Pléistocène moyen de La Micoque (Les Eyzies-de-Tayac, Dordogne) : *Equus mosbachensis micoquii nov. Ssp. Paléo- Revue d'archéologie Préhistorique*, 17, 73-110.
- L'HOMME, Marie-Claude (2004a) : *La terminologie : principes et techniques*. Montréal : Presses de l'Université de Montréal.
- L'HOMME, Marie-Claude (2004b) : Adjectifs dérivés sémantiques (ADS) dans la structuration des terminologies. In : *Actes du colloque Terminologie, ontologie et représentation des connaissances*. (Terminologie, ontologie et représentation des connaissances, Université Jean Moulin, Lyon, 22-24 juin 2004).
- L'HOMME, Marie-Claude et JIA, Zhichao (2015) : Classement des combinaisons lexicales spécialisées à base nominale dans un dictionnaire d'informatique. In : *Cahiers de lexicologie*, 106, 229-251.
- LERAT, Pierre (2009) : La combinatoire des termes. Exemple : nectar de fruits. In : *Hermès Journal of Language and Communication Studies*, 42, 211-232.
- MANIEZ, François (2002) : Distinguer les termes des collocations : étude sur corpus du patron <Adjectif-Nom> en anglais médical. In : *Actes du colloque Taln'02*. (Taln'02, Atilf et Loria, Nancy, 24-27 juin 2002).
- MAZIERE Florent (2001) : L'occupation des sols dans la moyenne vallée de l'Orb à la fin de l'âge du Bronze. *Documents d'archéologie médiévale*. 24, 88-115

- MENDES, Sara (2003) : Sémantique de l'adjectif et ses restrictions de sélection. *In : Actes EDIT*. (EDIT, Université Paul Sabatier, Toulouse, 2003)
- MIEYAA Yoan, ROUYER Véronique et LE BLANC Alexis (2012) : La socialisation de genre et l'émergence des inégalités à l'école maternelle : le rôle de l'identité sexuée dans l'expérience scolaire des filles et des garçons. *L'orientation scolaire et professionnelle*. 41(1).
- MONDADA Lorenza et TRAVERSO Véronique (2005) : (Dés) alignement en clôture. Une étude interactionnelle de corpus de français parlé en interaction. *LIDIL*. 31, 35-39.
- MONCEAUX, Anne (1997) : Adjectifs de relation, complémentation et sous-classification. *In : NAM Jee-Sun dir. La description syntaxique des adjectifs pour les traitements informatiques, Langages*, vol. 31-3, 39-59.
- ROCHETEAU, Jérôme et DAILLE, Béatrice (2011) : TTC TermSuite : a UIMA Application for Multilingual Terminology Extraction from comparable Corpora. *In : Processing of the 5th International Joint Conference on natural Language Processing (IJCNLP, Chiang Mai 8-13 November 2011)*.
- SCHMID, Helmut (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. *In : Proceedings of International Conference on New Methods in Language Processing*. Manchester.
- TOMA, Alice (2002) : L'interdisciplinarité et la terminologie mathématique : la migration des termes scientifiques, *In : Actes Glatt 2002*, Evry, 251-262.
- TUTIN, Agnès (2007) : Modélisation et annotations des collocations : une application au lexique transdisciplinaire des écrits scientifiques. *In : Formaliser les langues avec l'ordinateur : de INTEX à Nooj. Les cahiers de la MSHE – Archive, bases, corpus*. Besançon : Presses universitaire de Franche Comté, vol. 8.
- TUTIN, Agnès et GROSMANN, Francis (2014) : *L'écrit scientifique : du lexique au discours*, Rennes : Presses universitaires de Rennes. Collection Rivages linguistiques, 246p.
- VALLEE Nathalie, ROUSSET Isabelle et BOE Jean-Louis (2001) : Des lexiques aux syllabes des langues du monde. *LINX*. 48 : 37-50.
- VALLIN Luc, MASSON Bertrand, CASPAR Jean-Paul et DEPIEREUX Eric (2006) : L'outil idéal. Analyse du standard Levallois des sites moustériens d'Hermies (Nord de la France). *Paléo*. 16, 237-272.
- VANHAEREN Marian et D'ERRICO Francesco (2001) : La parure de l'enfant de la Madeleine (fouille Peyrony). Un nouveau regard sur l'enfance au Paléolithique supérieur. *Paléo*, 13, 201-240.
- WUSTER, Eugen (1981) : L'étude scientifique générale de la terminologie, zone frontière entre la linguistique, la logique, l'ontologie l'informatique et les sciences des choses. *In : Vladimir Ivanovich Sikorov (dir), Textes choisis de terminologie. I. Fondements théoriques de la terminologie*, Laval : Girsterm, 57-114.