



Disponible en ligne sur www.sciencedirect.com



ScienceDirect

et également disponible sur www.em-consulte.com



Article original

La version française du dictionnaire pour le LIWC : modalités de construction et exemples d'utilisation

The French dictionary for LIWC: Modalities of construction and examples of use

A. Piolat^{a,*}, R.J. Booth^b, C.K. Chung^c, M. Davids^d, J.W. Pennebaker^c

^a Centre PsyCLÉ, Aix-Marseille université, 29, avenue R.-Schuman, 13621 Aix-en-Provence cedex, France

^b University of Auckland, New Zealand, Australie

^c The University of Texas, Austin, États-Unis

^d Columbia University, New York, États-Unis

INFORMATIONS

Historique de l'article :

Reçu le 23 février 2011

Accepté le 20 juillet 2011

Mots clés :

LIWC

Comptage de mots

Analyse de textes

Traduction

Version française

RÉSUMÉ

LIWC 2007 est, à l'origine, un programme d'analyse qui compte les mots d'un texte de langue anglaise en les attribuant à des catégories qui ont un sens psychologique (<http://www.liwc.net/>). Il offre une analyse (en pourcentage) de 80 dimensions du langage (mots fonctionnels, thèmes, ponctuation). L'objectif de cette note méthodologique est de présenter la version française du logiciel LIWC. Nous présentons la structure des dimensions de la version anglaise du logiciel tout en donnant quelques éléments du mode d'emploi comme la préparation des textes à analyser. Nous expliquons, ensuite, les décisions prises pour opérer la traduction du dictionnaire anglais en français. Nous insistons sur les contraintes imposées par la morphologie du français écrit et les difficultés que nous avons rencontrées (élision de l'article et de la négation, temps des verbes). La présentation des problèmes de traduction permet de comprendre les modalités de construction du dictionnaire et permet donc à un utilisateur du LIWC de procéder à la construction d'un dictionnaire personnalisé qu'il lui permettra d'envisager, si nécessaire, l'analyse de contenus plus ajustés à ses recherches.

* Auteur correspondant.

Adresse e-mail : annie.piolat@univ-provence.fr (A. Piolat).

Une comparaison qualitative des indicateurs obtenus avec les versions anglaise et française du dictionnaire pour 66 textes bilingues de genre et de contenu très varié donne des résultats satisfaisants. Une comparaison statistique de 119 écritures expressives produites par des étudiants de trois cursus universitaires (lettres, sciences et psychologie) sur un événement (échec ou réussite à un examen) montre la validité de la version française pour identifier les contenus exprimés. Ainsi, l'outil mis à disposition sera efficace pour entreprendre des recherches dans les domaines psychologiques de la santé, du travail et de l'éducation à propos du langage (oral et écrit) produit dans différents contextes.

© 2011 Société française de psychologie. Publié par Elsevier Masson SAS. Tous droits réservés.

A B S T R A C T

Keywords:

LIWC

Word count

Text analysis

Translation

French dictionary

LIWC is, originally, a text analysis program that counts words of English texts in psychologically meaningful categories. It provides an analysis (in percentage) for 80 dimensions of language (functional words, topics, punctuation). The goal of this methodological note is to present the French LIWC. This version respects the structure of the categories of the English version of the software and gives explanations about the user guide such the preparation of the to be analyzed texts. Then, we explain the decisions for translating the English dictionary into French. We emphasize the constraints imposed by the morphology of written French language and the difficulties encountered (elision of the article, elision of negation and of verbs' tense). Presenting the translation problems allows understanding the modalities of construction of the dictionary and allows for a LIWC user to build a personalized dictionary in order to analyze contents more suited to the research needs. A qualitative comparison of dimensions obtained with both the French and English versions for 66 bilingual texts of various types and contents provides satisfactory results. A statistical comparison of 119 expressive writings produced by students from three university courses (Humanities, Sciences and Psychology) about a given event (success or failure to an exam) shows the validity of the French version for identifying the expressed contents. Thus, this tool should be efficient for undertaking research in different fields of psychology (health, work and education) concerning oral and written language produced in different contexts.

© 2011 Société française de psychologie. Published by Elsevier Masson SAS. All rights reserved.

1. Introduction

La version actuelle du logiciel «*Linguistic Inquiry and Word Count*» (LIWC) a été conçue par Pennebaker et al. (2007). Elle est disponible sur le site : <http://www.liwc.net/>. Ce logiciel analyse automatiquement le contenu de textes et calcule comment des personnes utilisent différentes catégories des mots qu'elles ont insérées dans leurs productions orales ou écrites (emails, discours, poèmes, dialogues quotidiens transcrits, etc.). Il permet de déterminer la proportion (en pourcentage) selon laquelle ces personnes utilisent des termes qui dénotent, par exemple, des émotions positives ou négatives, ainsi que des termes dont le contenu qui réfère à la sexualité, la maladie, l'alimentation ou la vie religieuse. Techniquement, ce logiciel identifie en quelques secondes plus de 80 dimensions du langage au sein d'un très grand nombre de textes donnés simultanément à analyser. Les données

obtenues sont facilement transférables dans des fichiers propices à une exploitation statistique (ex. : Excel, SPSS).

Les travaux de Pennebaker et de son équipe sont très renommés depuis une trentaine d'années. Aussi, les chercheurs qui s'inscrivent dans cette lignée d'études ont-ils procédé à la traduction du dictionnaire dans d'autres langues afin de pouvoir réaliser des études comparables. La traduction en espagnol a été l'une des premières entreprises (Ramirez-Esparza et al., 2007) et a permis par exemple une publication renommée sur les différences d'activités verbales selon le genre (Mehl et al., 2007). Depuis moins d'une dizaine d'années, plusieurs versions ont été élaborées : version allemande, arabe, chinoise, hollandaise, italienne, turque et russe. Cette note méthodologique annonce la création de la version française.

Dans le cadre de cet article, après avoir évoqué très succinctement le rôle que joue l'outil pour identifier les pratiques langagières des individus selon des déterminants internes (ex. : dépression) et externes (ex. : genre), nous allons présenter rapidement un des éléments fondamentaux du LIWC : la structure de son dictionnaire anglais. Notre travail de transposition en français n'a concerné que ce dictionnaire. Puis, dans la deuxième section, nous aborderons la procédure et les décisions que nous avons prises lors de cette transposition. Dans la troisième partie, nous opérerons une comparaison qualitative entre des textes bilingues anglais et français selon les indices calculés par le LIWC. Puis enfin, nous exposerons quelques résultats que l'on peut obtenir en appliquant la version française à une centaine de productions écrites estudiantines issues d'une précédente recherche (Piolat et Bannour, 2010). Nous concluons sur l'urgence qu'il y a à entreprendre des observations et des expérimentations enfin réalisables, grâce à la version française du LIWC, auprès d'individus francophones qu'il s'agisse d'identifier des spécificités langagières de tels ou tels groupes sociaux évoluant dans des contextes particuliers ou bien d'analyser l'impact de ce qui est écrit (ou dit) sur la santé physique et psychologique des personnes (pour une explication de ces champs thématiques de recherches fondamentales ou appliquées, voir la méta-analyse de Frattaroli, 2006 ainsi que la synthèse de Tausczik et Pennebaker, 2010).

2. Le *Linguistic Inquiry and Word Count* 2007 : fonction, structure du dictionnaire anglais et quelques conseils d'utilisation

Cette section a comme objectif de présenter succinctement des informations qui ont été largement présentées par Pennebaker et son équipe (Pennebaker et al., 2007 ; Tausczik et Pennebaker, 2010). Nous les évoquons ici afin que le lecteur comprenne plus aisément la nature de la version française du LIWC, version qui reproduit le mode de catégorisation de fonctions grammaticales et du sens inclus dans la version anglaise du LIWC.

2.1. Un logiciel au service de l'analyse sociale et psychologique du sens des mots

La version actuellement disponible en langue anglaise du LIWC 2007 (Windows, v1.11 ; Macintosh, v1.13) est le résultat d'une longue mise au point que Pennebaker et al. (2007) ont décrite dans leur présentation des modifications apportées à ce logiciel. Les objectifs de cette équipe étaient de disposer d'un outil automatique de détection et de comptabilisation des mots dans des textes qui évite la subjectivité des analyses de contenu opérées par des juges. Leur pari méthodologique était que, face à la complexité des procédés linguistiques qui permettent l'expression des ressentis, l'analyse du seul lexique produit associée au décompte des certaines catégories grammaticales (ex. : les pronoms personnels) constitue une démarche incomplète, mais efficace. La première version a été initialement développée pour explorer le langage de confessions écrites concernant des événements négatifs de vie et déterminer quelles caractéristiques langagières permettraient de prédire l'amélioration de la santé physique et psychique (Pennebaker, 1993 ; Pennebaker et al., 1997). Puis, ce logiciel a été perfectionné afin de pouvoir réaliser l'analyse de différentes productions écrites, mais aussi orales comme des textes littéraires, des essais personnels, des conférences de presse, des blogs, des conversations de la vie quotidienne, des réponses à des questions ouvertes (Pennebaker et Graybeal, 2001). Chemin faisant, les deuxième (LIWC 2001) et troisième (LIWC 2006) versions ont bénéficié d'un dictionnaire plus étendu et d'une interface homme-machine plus performante. Certaines des options d'analyses ont

été améliorées (ex. : segmentation du texte, choix de catégories d'analyse prédéfinies par l'utilisateur). Enfin, la version actuelle (LIWC 2007) permet de construire et d'exploiter un dictionnaire personnalisé pour des fins de recherche spécifiques.

Dans une synthèse très complète, Tausczik et Pennebaker (2010) ont rassemblé les résultats issus de 121 études qui ont exploité le logiciel LIWC. Leur objectif majeur est de montrer la façon dont certaines dimensions langagières que le LIWC identifie sont corrélées avec les caractéristiques suivantes : focus intentionnel des locuteurs (qui parle et à propos de quelle temporalité de l'événement), états émotionnels positifs et négatifs, types de relations interpersonnelles et les effets du statut de la dominance et de la hiérarchie sociale, modalités de coordination sociale et les effets de groupe, honnêteté et tromperie, styles de pensée, différences interindividuelles. Ainsi, selon ces auteurs, les mots utilisés quotidiennement reflètent ce que nous sommes et les relations sociales que nous établissons. La mise au point de cet outil a permis aussi de nombreuses études concernant l'impact bénéfique de l'écriture et particulièrement de l'écriture expressive sur les évolutions de la santé physique et psychologique des individus (pour des synthèses : Frattaroli, 2006 ; Piolat et Bannour, 2011 ; Smyth et Pennebaker, 2008). L'éventuel utilisateur du LIWC se doit d'accéder à ce large éventail de travaux afin d'identifier les types de dimensions linguistiques, donc de catégories du dictionnaire présentées ci-après, qu'il étudiera avec intérêt. Les catégories du dictionnaire anglais sont présentées dans la section qui suit. Les identifier lui permettra de repérer celles qui lui seront plus spécifiquement utiles pour analyser les textes (produits oralement ou par écrit) qu'il compte soumettre à l'analyse. Enfin, lors de l'acquisition du dictionnaire (<http://www.liwc.net/>), l'utilisateur pourra choisir d'acheter une ou plusieurs langues si son activité scientifique l'invite à des comparaisons inter-langues : anglais, arabe, chinois, hollandais, français, allemand, italien, russe, espagnol, turc.

2.2. La structure du dictionnaire anglais et français

Une refonte complète de la structure du dictionnaire anglais a été entreprise (nombre de mots, hiérarchie des catégories) en 2007. Afin d'y parvenir, Pennebaker et al. (2007) ont exploité plusieurs milliers de textes cumulés à l'issue de plus de 15 ans de recherche. Après avoir constaté que certains mots fonctionnels étaient étroitement liés à certains processus psychologiques, plusieurs catégories grammaticales ont été rajoutées (conjonctions, adverbes, quantifieurs, verbes et verbes auxiliaires, pronoms impersonnels, etc. ; Chung et Pennebaker, 2008). En revanche, d'autres catégories thématiques rarement utilisées par les rédacteurs et les locuteurs ont été supprimées. La structure du dictionnaire comporte cinq macro-catégories, elles-mêmes subdivisées en sous-catégories. Nous la présentons ci-après, car nous l'avons gardée pour la version française du LIWC (Tableau 1). Les effectifs qui sont présentés dans le Tableau 1 concernent les effectifs du LIWC 2007 anglais. Les effectifs du LIWC français sont très légèrement supérieurs pour des raisons présentées au début dans la section 3.1. (traduction dans quelques cas d'un mot anglais par deux mots français), mais nous n'avons pas, à ce jour, procédé au décompte exact de ces effectifs en français. La collection de mots pourrait être largement augmentée comme le montreront les résultats que nous présentons dans la section 4 en comparant des textes bilingues « tout-venant ». À titre d'exemple, la collection des mots relatifs aux émotions dans le dictionnaire anglais est de 915 termes qui sont : (émotions positives = 406 et émotions négatives = 499). Parmi les émotions négatives, 91 radicaux sont classés dans la sous-catégorie « anxiété », 184 dans la sous-catégorie « colère » et 101 dans la sous-catégorie « tristesse ». Compte tenu d'un travail précédent (Piolat et Bannour, 2009a), nous savons qu'une fois traduit en français, cet échantillon de termes anglais n'épuise pas les termes qui sont disponibles pour rendre compte des états affectifs et émotionnels (recensement de 2299 termes de sens propre). Leur mode de catégorisation au sein de chaque valence (mots positifs et mots négatifs) pourrait, aussi, être complété (cf. la structure du lexique d'EMOTAIX-Tropes ; Piolat et Bannour, 2009a).

2.3. Quelques indications concernant l'application du LIWC aux corpus à analyser

Dans le mode d'emploi fourni avec le logiciel, d'indispensables indications sont données afin de lancer le programme et l'exploiter (gestion : des dictionnaires, des fichiers de textes, de la sauvegarde des données sous forme de tableaux Excel ou autre, etc.). D'autres indications concernent la nécessaire

Tableau 1
Les 80 descripteurs analysés par le LIWC 2007 version anglaise (extrait de Pennebaker et al., 2007 ; NB : entre parenthèses l'effectif de radicaux présents dans le dictionnaire anglais).

Processus linguistiques	Processus psychologiques	Préoccupations personnelles	Dimensions du langage oral	Ponctuation
Total de mots	Processus sociaux (465)	Travail (327)	Consentement (30)	Total
Mots par phrase	Famille (64)	Accomplissement (186)	Phatiques (8)	Points
Mots du dictionnaire	Amis (37)	Loisirs (229)	Remplisseurs (9)	Virgules
Mots de plus de 6 lettres	Humains (61)	Maison (93)		Doubles points
Total de mots fonctionnels (464)	Processus affectifs (915)	Argent (173)		Points virgules
Total des pronoms (116)	Émotions positives (406)	Religion (159)		Points d'interrogation
Pronoms personnels (70)	Émotions négatives (499)	Mort (62)		Points d'exclamation
1 ^{er} personne du singulier (12)	Anxiété (91)			Tirets
1 ^{er} personne du pluriel (12)	Colère (184)			Guillemets
2 ^e personne (20)	Tristesse (101)			Apostrophes
3 ^e personne du singulier (17)	Processus cognitifs (730)			Parenthèses
3 ^e personne du pluriel (10)	Perspicacité (195)			Autres ponctuations
Pronoms impersonnels (46)	Causation (108)			
Articles (3)	Divergence (76)			
Verbes (383)	Tentative (155)			
Verbes auxiliaires (144)	Certitude (83)			
Verbes au passé (145)	Inhibition (111)			
Verbes au présent (169)	Inclusion (18)			
Verbes au futur (48)	Exclusion (17)			
Adverbes (69)	Processus perceptifs (273)			
Prépositions (60)	Vue (72)			
Conjonctions (28)	Audition (51)			
Négations (57)	Toucher (75)			
Quantificateurs (89)	Processus biologiques (567)			
Nombres (34)	Corps (180)			
Jurons (53)	Santé (236)			
	Sexualité (96)			
	Alimentation (111)			
	Relativité (638)			
	Mouvement (168)			
	Espace (220)			
	Temps (239)			

préparation des textes qui seront soumis à l'analyse. Nous en présentons quelques-unes ici, car elles permettent de comprendre certains des principes qui guident la construction d'un dictionnaire piloté par le logiciel.

2.3.1. Préparation des textes

Pour faciliter le travail du logiciel et s'assurer d'obtenir les meilleurs résultats, il est indispensable de bien préparer les textes avant de lancer l'analyse. Il faut procéder à un contrôle intransigeant de l'orthographe dite d'usage (ex. : dans les corpus, le mot « dégoût » ne doit pas être écrit de la façon suivante « dégout » sinon il ne sera pas comptabilisé), comme de celle dite grammaticale (justesse : des terminaisons des conjugaisons verbales, des accords du participe passé, etc.). Toutefois, dans la mesure où le LIWC2007 convertit tous les fichiers texte en minuscules avant de les traiter, il n'est pas nécessaire de corriger l'emploi des majuscules ainsi que la structure des phrases, mais la ponctuation finale des phrases doit être présente. Les abréviations doivent être remplacées par les mots qu'elles désignent, si l'on souhaite que le logiciel puisse les identifier et les comptabiliser (« janv » deviendra « janvier »). Les sigles comme « SNCF », peuvent être conservés à moins de vouloir que soit décompté chaque mot qui compose le sigle. Non inclus dans le dictionnaire, les sigles ne seront pas identifiés et comptabilisés à moins de les y introduire volontairement. L'usage du trait d'union doit être pratiqué de façon circonspecte. Dès qu'il est présent, il lie les séquences de caractères qui le précèdent avec celles qui le suivent en un seul « mot ». Par exemple, « beaux parents » sera analysé comme deux mots différents, alors que le dictionnaire reconnaît « beaux-parents » dont l'orthographe est plus exacte quand il s'agit d'un lien de parenté. Pour les expressions avec un trait d'union comme « est-il . . . ? », il convient de changer « est-il » en « est il » afin que l'auxiliaire et le pronom soient analysés. En raison de l'indice « words per sentence » (WPS) [mots par phrase] qui implique le dénombrement des signes de fin de phrase (point, point d'interrogation, points d'exclamation), il est utile d'ôter le ou les points associés aux abréviations courantes (Dr., U.S.A.) et de ne garder que la ponctuation de fin de phrase. Enfin, certaines habitudes orthographiques liées à l'usage de médias particuliers comme les SMS et les emails doivent être transcrites en orthographe conventionnelle (ex. : @ => à, & => et ; NB dans le cas de recherches spécifiques, ces symboles ainsi que les sigles peuvent être introduits dans le dictionnaire ou faire l'objet d'un dictionnaire spécialisé construit à cette fin).

2.3.2. Analyse de productions écrites et orales

Le LIWC peut aussi être utilisé pour analyser des conversations ou des entretiens. Toutefois, ce dictionnaire n'est pas adapté pour mettre en évidence, par exemple, les différences de tournures syntaxiques entre l'oral et l'écrit. En effet, il faut bien noter que le logiciel ne peut analyser efficacement que des textes qui ont été « toilettés » comme indiqué précédemment. De plus, il est indispensable de mettre en forme la transcription de certaines dimensions du langage parlé. Les hésitations (catégorie 463) peuvent être codées. La question est de savoir combien de fois il est nécessaire de les répéter. Certains éléments de dialogues comme les interrogations (« hein ? ») doivent être transcrits selon le format qu'offre le dictionnaire. Les remplisseurs encore appelés phatiques (catégorie 464) et les formules stéréotypées comme « tu vois » doivent être transcrits sous forme d'une seule séquence (« tuvois »), car elles sont introduites sous ce format dans le dictionnaire. Les commentaires adossés qui ont pu être surajoutés lors de la transcription écrite des dialogues afin de qualifier certains traits suprasegmentaux ne doivent pas être maintenus (ex. : accompagnement par des rires, voix qui tremble ou qui chuchote). Quand ce que dit une personne est incompréhensible, il est possible de transcrire « xxx », qui indique l'absence d'audibilité des propos, mais le logiciel, qui ne contient pas cet élément, ne pourra pas le comptabiliser sauf en l'introduisant volontairement dans un dictionnaire personnalisé.

3. La transposition du dictionnaire anglais en français : procédure, décisions et constats

Cette section a comme premier objectif de rendre compte des étapes de traduction pour passer du dictionnaire anglais au dictionnaire français. Son second objectif est de souligner les décisions qu'il a fallu prendre pour coder certaines particularités morphologiques du français écrit. En explicitant ainsi la procédure et nos modalités de travail, nous pensons avertir les usagers des fonctionnalités du dictionnaire mis à leur disposition ainsi que les futurs constructeurs d'un dictionnaire personnalisé en

Tableau 2

Présentation du codage du dictionnaire français exporté du LIWC 2007 anglais.

1	Fonction	17	Préposition	131	Cognition	149	Sexualité
2	Pronom	18	Conjonction	132	Perspicacité	150	Alimentation
3	Pronomp	19	Négation	133	Cause	250	Relativité
4	Je	20	Quantifieur	134	Divergence	251	Mouvement
5	Nous	21	Nombre	135	Tentative	252	Espace
6	Vous/tu	22	Juron	136	Certitude	253	Temps
7	Il	121	Social	137	Inhibition	354	Travail
8	Ils	122	Famille	138	Inclusion	355	Accomplissement
9	Pronomimp	123	Ami	139	Exclusion	356	Loisir
10	Article	124	Humain	140	Perception	357	Maison
11	Verbe	125	Affect	141	Voir	358	Argent
12	Verbe auxi	126	Émopos	142	Entendre	359	Religion
13	Verbe passé	127	Émonég	143	Sentir	360	Mort
14	Verbe présent	128	Anxiété	146	Biologique	462	Consentement
15	Verbe futur	129	Colère	147	Corps	463	Hésitation
16	Adverbe	130	Tristesse	148	Santé	464	Remplisseur

langue française. Il s'agit, de fait, de décrire certaines difficultés de codage qu'il a fallu surmonter (NB : la construction d'un dictionnaire personnalisé est indiquée dans le mode d'emploi anglais du logiciel qui est téléchargeable lors de l'acquisition du logiciel).

3.1. La procédure de codage et ses conséquences sur la taille du dictionnaire

Dans le **Tableau 1** sont présentées de façon organisée les dimensions syntaxiques (ex. : les pronoms) et thématiques (ex. : les émotions) qui sont contenues dans le dictionnaire. Le dictionnaire LIWC 2007 est constitué, quant à lui, d'une liste de mots ordonnés alphabétiquement et suivi d'un ou plusieurs codes chiffrés qui renvoient aux dimensions affichées dans le **Tableau 1**. La liste du codage des dimensions est insérée au début du dictionnaire, obligatoirement ouverte et fermée par le symbole « % ». Ce codage permet d'attribuer à ces termes des fonctions syntaxiques (ex. : « 2 » pour les pronoms ; « 13 » pour un verbe conjugué au passé) ainsi que certains éléments de signification sous forme de renvoi à des catégories de contenu. Les modalités de codage du dictionnaire français, transposées fidèlement du dictionnaire anglais, sont présentées dans le **Tableau 2**. Concrètement, ces catégories sont combinées lorsque la fonction et le sens du terme l'imposent. Par exemple, dans un énoncé où il est dit : « les étudiants t'aimeront », le logiciel catégorisera la séquence de caractères ininterrompue « t'aimeront » en le rapprochant de : « t'aimero* 11 15 121 123 125 126 146 149 1 2 3 6 ». Dans cette identification, il est indiqué que c'est un verbe conjugué au futur : « 11 15 ». Sa dénotation est indiquée par les catégories suivantes : « 121 123 125 126 146 149 ». Il est aussi mentionné que le pronom élide qui lui est attaché est un pronom personnel de la deuxième personne (toi : « 1 2 3 6 »).

L'utilisateur du LIWC peut visualiser sans problème la liste des mots codés que contient le dictionnaire qu'il exploite. Dans le **Tableau 3**, nous présentons un très bref extrait du dictionnaire français et du dictionnaire anglais concernant le terme « amour ».

À ce jour, le dictionnaire anglais (version 2007) comporte 4496 lignes qui correspondent chacune à un terme anglais, soit sous forme de radical « lover* », soit sous forme complète « loveley ». Pour le dictionnaire français, le nombre de lignes est presque neuf fois plus important ($n = 39\,164$). Pour comprendre un tel écart, plusieurs explications méritent d'être données.

Cet écart provient tout d'abord de notre volonté d'ajouter quelques nouveaux mots dans le dictionnaire en raison de la transposition des termes anglais en français. Dans un premier temps, nous avons procédé à la traduction de la version anglaise du LIWC terme à terme. Quand plusieurs mots français étaient possibles, nous avons choisi le terme français le plus fréquent. Toutefois, dès qu'un autre de ces mots nous est apparu utile et fréquent, nous l'avons rajouté dans la base française. Dans un second temps, nous avons associé aux termes de la liste française, les catégories anglaises. Mais, nous avons systématiquement contrôlé si cette importation était valide. Pennebaker et al. (2007) ont catégorisé les termes selon leur sens le plus habituel et le plus stable, autrement dit selon leur dénotation. En cas de doute, nous avons contrôlé cette dénotation avec le dictionnaire en ligne CNRTL : <http://www.cnrtl.fr/>.

Tableau 3

Présentation du codage des dictionnaires français et anglais pour le terme « amour ».

amour	121	125	123	126	146	149		
amoureuse	121	125	123	126	146	149		
amoureuement	1	16	121	125	123	126	146	149
amoureuses	121	125	123	126	146	149		
amoureux	121	125	123	126	146	149		
amours	121	125	123	126	146	149		
d'amour	121	123	125	126	146	149		
d'amoureuse ^a	121	125	123	126	146	149	10	
d'amoureux	121	125	123	126	146	149	10	
d'amours	121	123	125	126	146	149		
l'amour	121	123	125	126	146	149		
l'amoureuse ^a	121	125	123	126	146	149	10	
l'amoureux	121	125	123	126	146	149	10	
<i>lovely</i>	125	126						
<i>lover^a</i>	121	123	125	126	146	149		

Rappel : le codage est disponible dans le [Tableau 2](#) (ex. : 121 = social, 125 = affect ; 126 = émopos.).

^a L'astérisque indique une flexion possible (cf. section 3.1).

La catégorisation appliquée ne renvoie donc pas aux autres sens (connotation[s]) que ce terme peut prendre selon les contextes particuliers dans lesquels il peut être employé. En contrôlant la catégorisation, nous n'avons pas hésité à opérer une modification du codage (choix des catégories chiffrées) pour un nombre non négligeable de mots afin de mieux rendre compte du sens qu'ils dénotent en français. Ainsi, compte tenu de cette façon de procéder, les termes ont été catégorisés en évitant largement leur éventuelle polysémie ainsi que leurs possibles sens dénotatifs, autrement dit en évitant systématiquement les sens métaphoriques qu'ils peuvent revêtir.

Dans le même temps, nous avons maintenu autant que faire se peut le regroupement des termes autour d'un même radical. Ce radical qui est suivi d'une étoile mise en exposant permet d'économiser des lignes dans le dictionnaire ([Tableau 3](#)). Mais ce regroupement est techniquement très contraignant. Il nous a fallu nous assurer qu'un radical n'allait pas entraîner une comptabilisation par le logiciel d'un terme plus long comportant pour partie la même séquence de caractères alors qu'il ne dénote pas le même sens. Comparativement au dictionnaire anglais, nous avons souhaité ne pas trop rassembler de termes sous un même radical. Il nous a semblé important, en effet, de pouvoir coder, en plus du sens, certaines des propriétés syntaxiques des mots. Par exemple, le dictionnaire anglais comporte 69 adverbes (ex. : *anywhere* 1 16 131 135 250 252), dont quelques-uns qui se terminent par le morphème « ly » (ex. : *apparently* 1 16 131 135). Les adverbes français terminés par le morphème « ment » sont très nombreux et très usités. Aussi, lorsqu'un adjectif fréquent pouvait être dérivé à partir d'un des termes présents dans le dictionnaire, nous lui avons volontairement préservé une ligne de codage ([Tableau 3](#) : « amoureuement »).

3.2. Le codage de certaines propriétés morphologiques du français : conséquences et limites

Le logiciel LIWC analyse des « mots », c'est-à-dire une séquence de caractères comprise entre deux absences de caractère, autrement dit deux « blancs ». Cette façon de procéder a rendu délicate la gestion de certaines des propriétés morphologiques du français qui, par exemple, opèrent des liaisons entre termes. Pour coder ces particularités morphologiques du français, nous avons dû introduire dans le dictionnaire un très grand nombre de lignes comme nous l'expliquons ci-après. On peut rappeler qu'en traduisant le dictionnaire LIWC 2001 de l'anglais vers l'espagnol, [Ramirez-Esparza et al. \(2007\)](#) ont rencontré des problèmes du même ordre que ceux que nous avons affrontés pour la transposition en français. Le nombre de lignes du dictionnaire espagnol est d'ailleurs égal à 12 621, soit trois fois plus que la version anglaise.

La première difficulté de codage a été provoquée par l'usage en français des clitiques et la façon de les orthographier. Un clitique est, en linguistique, un élément à mi-chemin entre un mot indépendant et un morphème lié. En français (et dans les langues romanes en général), les articles, les pronoms personnels sujets ou d'objet (direct ou indirect) ainsi que la préposition *de*, sont des clitiques. Ils

sont majoritairement écrits de façon isolée entre deux espaces. Le logiciel peut alors les comptabiliser comme tel (ex. : « je te donne un grand amour » ; « je te le donne »). Les unités « je », « te », « le » et « un » sont comptabilisées et catégorisées pour ce qu'elles sont (pronom personnel, article). L'unité « amour » est aussi comptabilisée en étant associée à ses descripteurs sémantiques à l'aide des catégories chiffrées. Mais, dans de nombreux cas, ces clitics sont élidés et transcrits orthographiquement de façon liée à l'aide de l'apostrophe (ex. : « je te porte beaucoup d'amour. Je t'aime »). Si les séquences avec élision ne sont pas introduites volontairement dans la liste des termes du dictionnaire, des séquences comme « d'amour » et « t'aime » ne peuvent pas être comptabilisées par le logiciel qui, repérant l'apostrophe comme un caractère, les considère comme un « mot ». Or, devant les termes dont l'orthographe débute par une voyelle (a, e, i, o, u, y) et même par « h », l'élision de l'article (l') y compris de l'article partitif (d'), d'un pronom personnel (j', t', m', s') est un phénomène extrêmement fréquent et orthographiquement obligatoire en français. De ce fait, pour tous les termes du dictionnaire français commençant par une voyelle, il a donc fallu entrer les séquences liées avec un article ou un pronom et opérer un codage concernant à la fois l'élément syntaxique (article, pronom), mais aussi l'élément de contenu (ex. : « m'émervillai* 11 13 125 126 1 2 3 4 »).

La deuxième difficulté de codage a été entraînée par un effet morphologique du même ordre : la négation « ne + verbe + pas ». Dès qu'un verbe commence par une voyelle, la particule « ne » doit être élidée et liée (ex. : « je n'aime pas »). Comme précédemment, et toujours pour rendre possible le décompte de la négation élidée comme celle du contenu du verbe qui lui est associé, nous avons dû insérer pour chacun des verbes concernés une ligne suivie du codage approprié (ex. : « n'encourages 11 14 121 125 126 19 » ; « 19 » renvoyant à la négation). Cette introduction a dû être opérée pour chacune des déclinaisons du verbe imposé par la conjugaison de ces verbes. Cette nécessité a provoqué, là encore, une démultiplication très conséquente de ligne de codage.

La troisième difficulté de codage qui a considérablement ralenti la mise au point de la version française du dictionnaire concerne la question du temps des verbes. La conjugaison anglaise est fort simple (y compris avec les verbes irréguliers) comparativement aux nombreuses flexions qui, en français, accompagnant le radical verbal et qui permettent d'exprimer les différences de personne, de nombre, de mode et de temps. Nous avons limité nos objectifs et fait en sorte de rendre compte des changements morphologiques imposés par le présent, le futur, l'imparfait et le passé simple. De plus, afin de ne pas décliner toutes les flexions de chacun des verbes (dont beaucoup sont irréguliers) en français, tout en opérant un codage économique, nous avons exploité le plus possible l'exposant étoilé. Dans le cas des quelques homographes morphologiques pour lesquels une même séquence orthographique peut être codée comme étant du présent ou du passé, nous n'avons pas voulu associer à la ligne un codage contradictoire en indiquant les deux temps. Nous avons opté pour le temps de verbe le plus fréquent pour un radical donné (ex. : « il dépérit » dont la forme est identique pour le présent ou le passé simple, nous avons préservé le présent : dépérit 11 14 125 127 146 148 360). De ce fait, le codage du temps des verbes que nous avons opéré à la suite de longs et fastidieux contrôles, reste, à nos yeux, le codage le plus insatisfaisant. Tous les temps n'ont pas été codés et, de plus, les temps qui imposent l'usage d'un auxiliaire (être et avoir) seront codés d'une façon impropre (ex. : « j'ai aimé » ne sera pas codé comme le passé composé du verbe aimer, mais chacun des mots sera codé (j'ai au présent et aimé au passé). Le chercheur qui voudra travailler spécifiquement sur le temps des verbes employés par des groupes participants en fonction de contextes particuliers devra compléter le dictionnaire français ou bien exploiter un logiciel plus efficace pour les détecter comme peut le faire par exemple le logiciel Tropes.

Au total, construire cette version du dictionnaire français piloté par le logiciel LIWC 2007 a demandé plus de trois ans. Tel quel, ce dictionnaire permettra d'identifier des caractéristiques des textes étudiés sans laisser place à la subjectivité de juges qui opère des analyses de contenu. Les potentialités de recherches seront nombreuses comme nous l'illustrerons rapidement dans la section 5. Toutefois, tel qu'il est, ce dictionnaire comporte aussi de nombreuses limites. Les analyses seront dépendantes des termes qui ont été introduits et qui, comparativement aux possibilités lexicales de dénotation de la langue française, seront pour certains types de contenu relativement peu discriminant. Pour les repérages des termes syntaxiques, surtout des certaines particularités syntaxiques comme le temps des verbes, il reste imprécis et incomplet. Enfin, le repérage du contenu des textes est prédéterminé par les catégories descriptives (Tableau 2). Selon la nature de la recherche, ces catégories peuvent s'avérer

suffisantes ou, au contraire, elles doivent être repensées et, en conséquence, impliquer la construction d'un dictionnaire personnalisé.

4. Comparaison de résultats obtenus à partir d'un échantillon de textes bilingues

Afin de comparer qualitativement la description des textes que les versions anglaise et française du LIWC 2007 nous offrent, nous avons cherché au hasard sur le web des textes bilingues de genres différents et de thématiques très contrastées. Sans nous préoccuper des compétences des traducteurs, ni de la qualité de la traduction d'une langue dans l'autre, nous avons récolté 66 textes : extraits de textes littéraires et de contes (ex. : Andersen, Proust), de poèmes et de textes de chansons (ex. : Baudelaire, Beatles), de recettes de cuisine (ex. : japonaise, française, muffins), de description de pays, d'institutions et de musées (ex. Québec, Bulgarie, Onu, Unesco, gendarmerie, Louvre, musée Grévin), d'essais politiques (ex. : Sarkozy, Obama, marche de protestation), d'informations encyclopédiques (ex. : révolution française, expérience scientifique, poussette), d'essais sur des sujets de société (ex. : mariage gay, site internet, fécondité, grippe, garde d'enfant, mode). Selon les conseils de Pennebaker, l'important était de disposer d'un échantillon vaste et hétérogène sur le plan stylistique et thématique.

Dans le [Tableau 4](#) qui rassemble les indices linguistiques fournis par le logiciel, nous constatons, tout d'abord, que l'ensemble des textes français comporte 19 599 mots et l'ensemble des textes anglais 18 266 mots.

Les dictionnaires actuels n'ont filtré que 27,78 % des termes français et 22,14 % des termes anglais présents dans ces deux échantillons de textes de genre et de contenu très variés. Ces faibles pourcentages mettent en évidence que pour réaliser des études de productions verbales concernant des thématiques très spécifiques et originales, le chercheur aura intérêt à étoffer le dictionnaire français qu'il utilisera. Toutefois, dans le cadre de la présentation des productions faites par des étudiants à propos des sentiments qu'ils ont ressentis lors de leur échec ou de leur réussite à un examen (cf. section 5), le dictionnaire français filtre beaucoup plus de mots (84 % en moyenne ; [Annexe 1](#)) et s'avère, ainsi, bien adapté aux études expérimentales qui consistent à comparer, par exemple, des écritures expressives (expression du ressenti émotionnel à propos d'un événement, d'un objet, d'une personne).

Le constat d'écart syntaxiques très importants entre le français et l'anglais (ex. : articles français = 17,6 % vs articles anglais = 7,70 %) n'est pas surprenant compte tenu des spécificités de chaque langue. En revanche, si, comme l'ont fait [Rude et al. \(2004\)](#), les chercheurs sont intéressés par l'usage des pronoms personnels « je » vs « nous » opérés par des locuteurs ou des rédacteurs, les pourcentages observés dans les deux langues sont du même ordre (je = 3,27 % vs I = 2,49 % ; nous = 0,73 % vs we = 0,83 %).

Tableau 4

Indicateurs syntaxiques associés à la totalité des textes anglais et français (en pourcentage sauf pour le nombre total de mots).

		Processus linguistiques								
Catégories	Segment	Tot. mots	Mots/phr.	Dictionn.	Mots 6 let.					
Français	1,00	19599,00	34,09	27,78	74,49					
Anglais	1,00	18266,00	30,75	22,14	79,88					
Catégories	Fonction	Pronom	Pronomp	Je	Nous	Vous	Il	Ils	Pro. Imp	Article
Français	46,27	11,44	11,10	3,27	0,73	1,70	4,86	0,59	0,34	17,62
Anglais	50,48	12,30	7,69	2,49	0,83	2,29	1,42	0,67	4,61	7,70
Catégories	Verbe	Verbe auxi.	Passé	Présent	Futur	Adverbe	Préposit.	Conjonct.	Négation	
Français	12,74	3,34	3,91	6,13	0,58	4,39	13,17	6,46	2,03	
Anglais	11,43	6,72	2,70	6,65	1,06	3,27	13,51	5,93	1,30	
Catégories	Quantifieur	Nombre	Juron							
Français	6,36	2,39	0,15							
Anglais	2,38	0,70	0,11							

Remarque : les pourcentages sont calculés par le logiciel. Il s'agit d'un ratio (nombre de fois ou des termes de la catégorie concernée apparaissent divisé par le nombre total de mots).

Tableau 5

Indicateurs thématiques associées à la totalité des textes anglais et français (en pourcentage).

Processus affectifs									
Catégories	Affect	Mopos	Monég	Anxié.	Colèr.	Trist.			
Français	7,00	4,94	1,69	0,31	0,52	0,48			
Anglais	5,34	4,01	1,29	0,21	0,41	0,30			
Préoccupations personnelles									
Catégories	Travail	Accomplis.	Loisir	Maison	Argent	Religion	Mort		
Français	2,19	2,65	1,32	0,54	1,19	0,52	0,10		
Anglais	1,89	1,92	1,20	0,47	0,72	0,55	0,11		
Processus biologiques									
Catégories	Perception	Voir	Entendre	Sentir	Biologique	Corps	Santé	Sexualité	Aliment.
Français	2,94	1,51	0,55	0,58	2,81	0,76	0,82	0,91	0,41
Anglais	2,38	1,07	0,60	0,49	2,51	0,74	0,65	0,82	0,38

Remarque : les pourcentages sont calculés par le logiciel. Il s'agit d'un ratio (nombre de fois ou des termes de la catégorie concernée apparaissent divisé par le nombre total de mots).

Dans le **Tableau 5**, nous présentons quelques-uns des indicateurs (processus affectifs, préoccupations personnelles, processus biologiques) obtenus pour les deux lots de textes. Le dictionnaire anglais et le dictionnaire français donnent une image très analogue des divers thèmes qui ont été exploités dans les deux échantillons de textes bilingues (NB : compte tenu des choix de transposition que nous avons faits et présentés dans la section 3, nous n'avons pas jugé utile de comparer statistiquement, comme Ramirez-Esparza et al., 2007, l'ont fait, les écarts ou les ressemblances entre les analyses réalisées à l'aide des dictionnaires anglais et français sur des textes bilingues).

5. Application de l'outil French LIWC à de textes expérimentaux

Afin de tester la capacité du dictionnaire français à spécifier certaines caractéristiques syntaxiques et de contenu du langage produit, nous avons appliqué l'analyse à des textes rédigés par 119 étudiants dans un contexte expérimental (Piolat et Bannour, 2010). Il n'est pas utile ici de préciser les raisons de la construction du design expérimental qui ont présidé à la réalisation de la recherche publiée en 2010. L'intérêt de ces corpus est qu'ils ont été produits dans des contextes différents. Ainsi, les participants étaient inscrits en première année de trois cursus universitaires : un cursus littéraire (lettres modernes et classiques), un cursus scientifique (mathématique, informatique ou sciences physiques) et un cursus de sciences humaines (psychologie). De façon aléatoire, ces étudiants ont été invités à s'exprimer par écrit à propos d'un des deux thèmes suivants :

- exprimez par écrit ce que vous avez ressenti lorsque vous avez échoué à un examen (BAC, permis de conduire, ou autre examen important pour vous). Détaillez vos émotions, vos sentiments ;
- exprimez par écrit ce que vous avez ressenti lorsque vous avez obtenu votre BAC (ou un autre examen important). Détaillez vos émotions, vos sentiments.

Ces thèmes ont été retenus car ils permettaient de contraster l'expression d'émotions positives (réussir à un examen) et l'expression d'émotions négatives (échouer à un examen). La répartition des effectifs d'étudiants en fonction des conditions expérimentales est présentée dans le **Tableau 6**.

Nous avons analysé les écrits avec le logiciel LIWC 2007 en le laissant calculer les pourcentages de toutes les catégories (fonction « *Use all categories* » ; le décompte fait à propos de la ponctuation ne sera pas évoqué ici). Puis, nous avons évalué statistiquement l'impact des deux facteurs expérimentaux (thématique de la production écrite et cursus universitaire) sur chacune de ces catégories à l'aide d'une Anova. Afin de ne pas allonger cette note méthodologique, les résultats sont regroupés dans l'**Annexe 1**.

Ces résultats peuvent être succinctement commentés de la façon suivante. Comme l'on pouvait s'y attendre, l'effet de la thématique de l'écriture expressive (réussite vs échec) est massif sur

Tableau 6

Effectifs des participant en fonction de la thématique de l'écriture expressive (échec vs réussite à un examen) et en fonction de leur cursus universitaire (lettres vs sciences vs psychologie).

	Lettres	Sciences	Psychologie	Ensemble
Échec	18	20	25	63
Réussite	16	16	24	56
Ensemble	34	36	49	119

certaines des indicateurs syntaxiques (ex., mots de six lettres, temps des verbes, négation). Dans le cadre d'une étude approfondie sur l'impact des examens, il serait ainsi intéressant de comprendre pourquoi les étudiants exploitent significativement plus le temps futur quand ils présentent un échec (frein à leur ambition, à un futur travail ?) et plus le temps passé quand ils présentent leur réussite (ambition accomplie et plaisir associé déjà loin ?). L'impact de l'événement (échec vs réussite) est très important sur quasiment toutes les catégories qui décrivent. L'engagement intellectuel (ex. : cognition), le ressenti affectif (ex. : termes émotionnels), la contrainte physique (ex. : biologie, santé) et, à un moindre degré, la relation aux autres (ex. : famille, amis) sont associés à la description de l'événement positif comme de l'événement négatif. De façon plus étonnante, la catégorie « accomplissement » est plus fréquemment abordée dans la situation d'échec que dans celle de la réussite à un examen. En associant une plus forte proportion de « négation » et de verbes au temps futur, les étudiants semblent traduire ainsi les conséquences importantes pour eux de la déconvenue associée à un insuccès.

De plus, les résultats statistiques regroupés dans l'[Annexe 1](#) mettent en évidence un impact moins fort du type de cursus sur ce qu'expriment les étudiants. L'analyse des effets principaux du facteur cursus et interactions significatives met en évidence que les étudiants qui suivent un cursus scientifique et qui relatent un échec ont des productions écrites qui se différencient de celles des autres étudiants sur plusieurs catégories qu'il s'agisse du volume verbal, de catégories syntaxiques (ex. : pronoms personnels, négation) ou de catégories de contenu (ex. : affect, loisir). Un échec à un examen détruirait leur ambition d'études exceptionnelles comme l'entrée dans une grande école. Les étudiants de psychologie évoquent plus, quant à eux, des questions d'argent et de travail (investissement dans le travail intellectuel pour réussir et nécessité pécuniaire d'avoir un petit job).

Ainsi, ces quelques éléments statistiques succinctement présentés témoignent de la capacité du LIWC et du dictionnaire français à rendre compte des différences de propos impliqués par la description d'un événement important : réussir ou échouer à un examen. Ils permettent aussi d'envisager l'intérêt qu'il y aurait à analyser les variations des propos d'étudiants de cursus différents. Selon leur cursus, les étudiants semblent vivre des expériences cognitives et affectives très différentes et les événements qui encouragent ou contrecarrent leurs objectifs ne paraissent pas avoir le même impact sur ce qu'ils évoquent de leur vécu. D'ailleurs l'évolution de leur anxiété avant et après avoir écrit, c'est-à-dire avant et après avoir évoqué à nouveau ces événements ne sont pas du même ordre selon que l'événement est positif (réussite) ou négatif (échec) et selon leur cursus comme nous l'avons montré par ailleurs ([Piolat et Bannour, 2009b, 2010](#)).

6. Conclusion

La célébrité du logiciel d'analyse de textes LIWC ([Pennebaker et al., 2007](#)) n'est plus à faire. [Smyth et Pennebaker \(2008\)](#) ou encore [Tausczik et Pennebaker \(2010\)](#) ont regroupé les types de recherche qu'il est possible d'entreprendre avec cet outil. Depuis 2007, les chercheurs ont souhaité pouvoir réaliser ces études dans d'autres langues que l'anglais et à ce jour plusieurs versions du dictionnaire sont disponibles (cf. <http://www.liwc.net/> : arabe, chinois, hollandais, anglais, allemand, italien, russe, espagnol, turc). La mise au point du dictionnaire français, composant indispensable pour que le logiciel puisse procéder à la quantification des catégories fonctionnelles et de sens contenu dans textes, a été à bien des égards délicate à entreprendre. Nous en avons donné les raisons dans cet article et, chemin faisant, nous avons souligné les avantages et les inconvénients associés à cette transposition. En utilisant des productions écrites d'une centaine d'étudiants, nous avons montré la forte capacité

discriminative de la version française du dictionnaire pour comparer des écritures expressives relatant un événement positif ou négatif.

Avec cet outil, il est maintenant possible de procéder, comme cela a été et continue d'être fait, à des recherches dans les différentes sous-disciplines de la psychologie qui posent des questions de santé, de travail et d'éducation et qui exploitent les dire et les écrits des personnes concernées par les études. À titre d'exemple, il sera envisageable de travailler sur des différences d'expression verbale entre les hommes et les femmes ou encore sur les effets des modalités d'expression orale et écrite (Augustine et al., 2011 ; Mehl et al., 2007). En psychologie de la santé et en psychopathologie, il sera enfin possible d'identifier dans les entretiens ou les écrits de patients francophones certaines caractéristiques linguistiques et thématiques de leurs « dire » (Rude et al., 2004). Les évolutions de leur état affectif et cognitif en fonction de thérapie comme, par exemple, la cure écrite pourront être mesurées. Il sera aussi envisageable de repérer si des étudiants francophones peuvent bénéficier d'une brève écriture expressive juste avant leur examen afin de diminuer leur anxiété et d'obtenir de meilleurs scores (Ramirez et Beilock, 2011). Par ailleurs, dans le cadre des études sur le partage social des émotions, les chercheurs pourront analyser les informations qui sont échangées par les différents groupes sociaux francophones (Rimé, 2009).

Enfin, le futur usager de la version française du dictionnaire LIWC pourra mettre au point un dictionnaire plus ajusté à ses besoins d'analyse. Le présent article en montre la démarche.

Déclaration d'intérêts

Les auteurs n'ont pas transmis de déclaration de conflits d'intérêts.

Remerciements

De nombreux contrôles fastidieux ont été entrepris pour parvenir à la version française du LIWC 2011. Les auteurs tiennent à remercier chaleureusement pour leur contribution persévérante et efficace Marlène Aillaud, Line Marie Béra, Rachid Bannour et Marion Phélip.

Annexe 1. Effets statistiques de la thématique (échec vs réussite) et du cursus (lettres, sciences et psychologie) sur quelques-uns des indicateurs thématiques (moyennes en pourcentage).

Indicateurs calculés par LIWC	Thèmes : E = échec ; R = réussite	Cursus : L = lettres ; S = sciences ; P = psychologie	Interaction
Mots du dictionnaire	Ns E = 83,9 ; R = 85,0	Ns L = 84,1 ; S = 84,9 ; P = 84,4	F(2,113) = 4,36*
Nbre de mots	Ns	F(1,113) = 10,14*** L = 271,1 > S = 193,7 < P = 237,3	F(2,113) = 5,19**
Mots > 6 lettres	F(1,113) = 6,46* E = 23,7 < R = 25,6	Ns	Ns
Vous + il/ils	Ns	F(1,113) = 8,43*** L = 6,6 < S = 8,1 > P = 6,5	F(2,113) = 4,62*
Verbes au passé	F(1,113) = 25,53*** E = 8,2 < R = 11,3	Ns	Ns
Verbes au présent	F(1,113) = 17,64*** E = 8,4 > R = 6,8	Ns L = 7,5 < S = 7,9 > P = 7,6	F(2,113) = 6,70**
Verbes au futur	F(1,113) = 5,06* E = 5,2 > R = 0,3	Ns	Ns
Négation	F(1,113) = 5,53* E = 4,0 > R = 3,2	F(1,113) = 3,62* L = 4,0 > S = 3,0 < P = 3,9	Ns
Famille	F(1,113) = 10,86*** E = 0,2 < R = 0,5	Ns	Ns

Annexe 1. (Suite)

Indicateurs calculés par LIWC	Thèmes : E = échec ; R = réussite	Cursus : L = lettres ; S = sciences ; P = psychologie	Interaction
Ami	F(1,113) = 11,70*** E = 0,2 < R = 0,4	Ns	Ns
Affect	Ns	Ns L = 11,3 < S = 12,7 > P = 11,3	F(2,113) = 3,55*
Mots positifs	F(1,113) = 59,21*** E = 4,2 < R = 7,1	Ns	F(2,113) = 4,57*
Mots négatifs	F(1,113) = 51,51*** E = 6,4 > R = 4,0	Ns	Ns
Anxiété	F(1,113) = 6,02* E = 0,8 < R = 1,2	Ns	Ns
Colère	F(1,113) = 8,90*** E = 0,9 > R = 0,4	Ns	Ns
Tristesse	F(1,113) = 77,46*** E = 3,1 > R = 1,2	Ns	Ns
Cognition	F(1,113) = 44,41*** E = 24,7 > R = 20,6	Ns	Ns
Biologie	F(1,113) = 24,65*** E = 2,9 > R = 1,9	Ns	Ns
Santé	F(1,113) = 45,41*** E = 2,3 > R = 1,2	Ns	Ns
Travail	F(1,113) = 40,48*** E = 4,6 > R = 2,8	Ns L = 3,9 ; S = 3,9 > P = 3,5	F(2,113) = 3,52*
Accomplissement	F(1,113) = 9,97** E = 5,5 > R = 4,5	Ns	Ns
Loisir	Ns	Ns L = 1,2 < S = 1,4 > P = 1,1	F(2,113) = 6,22***
Argent	Ns	F(1,113) = 5,64*** L = 0,5 ; S = 0,4 < P = 1,1	Ns

Ns = non significatif ; * $p < 0,05$; ** $p < 0,01$; *** $p < 0,005$

Références

- Augustine, A.A., Mehl, M.R., Larsen, R.J., 2011. A positivity bias in written and spoken English, and its moderation by personality and gender. *Social Psychology and Personality Science* 2 (5), 508–515.
- Chung, C.K., Pennebaker, J.W., 2008. Variations in spacing of expressive writing sessions. *British Journal of Health Psychology* 13, 15–21.
- Frattaroli, J., 2006. Experimental disclosure and its moderators: a meta-analysis. *Psychological Bulletin* 132, 823–865.
- Mehl, M.R., Vazire, S., Ramirez-Esparza, N., Slatcher, R.B., Pennebaker, J.W., 2007. Are women really more talkative than men? *Science* 317, 82.
- Pennebaker, J.W., 1993. Putting stress into words: health, linguistic, and therapeutic implications. *Behaviour Research and Therapy* 31, 539–548.
- Pennebaker, J.W., Graybeal, A., 2001. Patterns of natural language use: disclosure, personality, and social integration. *Current Directions* 10, 90–93.
- Pennebaker, J.W., Booth, R.J., Francis, M.E. (2007). *Linguistic Inquiry and Word Count (LIWC2007): A text analysis program*. Austin, TX: LIWC.net sur <http://www.liwc.net/>. (retrouvé le 11 juillet 2011).
- Pennebaker, J.W., Mayne, T.J., Francis, M.E., 1997. Linguistic predictors of adaptive bereavement. *Journal of Personality and Social Psychology* 72, 863–871.
- Piolat, A., Bannour, R., 2009a. EMOTAIX: un scénario de tropes pour l'identification automatisée du lexique émotionnel et affectif. *L'Année psychologique* 109, 657–700.
- Piolat, A., Bannour, R., 2009b. An example of text analysis software (EMOTAIX-Tropes) use: the influence of anxiety on expressive writing. *Current Psychology Letters. Brain, Behavior and Cognition* 5, 2, <http://cpl.revues.org/index4878.html>.
- Piolat, A., Bannour, R., 2010. Effets de l'expression écrite d'un événement positif et négatif sur le niveau d'anxiété d'étudiants de différentes disciplines. *Psychologie française* 55, 1–23.
- Piolat, A., Bannour, R., 2011. Les effets de l'écriture expressive sur la santé physique et psychologique des rédacteurs: un bilan, des perspectives de recherches. *Revue Européenne de Psychologie Appliquée* 61 (2), 101–113.
- Ramirez, G., Beilock, S.L., 2011. Writing about testing worries boosts exam performance in the classroom. *Science* 331, 211–213.
- Ramirez-Esparza, N., Pennebaker, J.W., Garcia, F.A., Suria, R., 2007. La psicología del uso de las palabras: un programa de computadora que analiza textos en Español (The psychology of word use: a computer program that analyzes texts in Spanish). *Revista Mexicana de Psicología* 24, 85–99.
- Rimé, B., 2009. Emotion elicits the social sharing of emotion: theory and empirical review. *Emotion Review* 1, 60–85.

- Rude, S.S., Gortner, E.M., Pennebaker, J.W., 2004. Language use of depressed and depression-vulnerable college students. *Cognition and Emotion* 18, 1121–1133.
- Smyth, J.M., Pennebaker, J.W., 2008. Exploring the boundary conditions of expressive writing: in search of the right recipe. *Revue Européenne de Psychologie Appliquée* 13, 1–7.
- Tausczik, Y., Pennebaker, J.W., 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29, 24–54.