Arob@se, www.arobase.to

volume 1-2, pp. 97-117, 2003

Codage et traitement automatique de corpus pour l'étude des prises de notes en français langue première et langue seconde

P Martine Faraco[1], Marie-Laure Barbier[2], Achille Falaise[3], Sonia Branca Rosoff[4]

1 Laboratoire Parole et Langage (UMR 6057, CNRS & UP), Université de Provence, Aix-en-Provence, France

2 Centre de Recherche en Psychologie de la Connaissance, du Langage et de l'Émotion (PsyCLÉ, EA 3273), Université de Provence, Aix-en-Provence, France

3 Laboratoire Linguistique et DIdactique des Langes Etrangères et Maternelles (LIDILEM, EA 609), Université Stendhal, Grenoble, France

4 Laboratoire Systèmes Linguistiques, Enonciation et Discursivité (SYLED, EA 2290), Université de Paris III-Sorbonne nouvelle, Paris, France

faraco@lpl.univ-aix.fr ; achillefalaise@hotmail.com ; ml.barbier@aix-mrs.iufm.fr ; branca@msh-paris.fr

Cette recherche a été réalisée dans le cadre du contrat AL 13b attribué par l'ACI « Ecole et Sciences Cognitives » que nous remercions.

Résumé

Afin de pouvoir procéder à des analyses quantitatives sur les procédures de prise de notes de scripteurs d'origine linguistique diverse, une analyse systématique des corpus doit être faite. Deux types de codages de prises de notes en langue première et en langue seconde sont proposés et décrits : (1) un codage traditionnel avec un tableur et (2) un codage automatisé en langage HML. La portée et les limites de chacun de ces codages sont analysées afin de poursuivre l'élaboration d'une grille d'analyse plus optimale pour rendre compte des procédés de prise de notes. Si les difficultés de comparaisons interlangues arrivaient à être résolues (l'exemple du japonais est proposé, ces grilles permettraient d'identifier les procédés transférables de la L1 à la L2, et de repérer ainsi les savoir faire mais aussi les difficultés rencontrées par les noteurs en français langue seconde.

Mots clés : Prise de notes, codage des informations, choix de critères, traitement automatique des données, L1/L2.

Abstract

Coding and automatic treatment of corpuses for studying note taking in French L1 and L2.

To pursue quantitative analysis of note-taker's procedures, from various native languages, a systematic analysis must be done. Two methods for coding data collected in L1 and L2 are suggested: (1) a traditional method as it is usually conducted in linguistic and psycholinguistic studies and (2) another one which allows the realization of automatic analysis by coding the data in HML language. The implications and limits of each of these methods are discussed in this chapter. The objective is to improve the analysis of note-taking procedures. Despite some constraints for inter-language comparisons which have to be solved (the example of Japanese note-taking is presented), automatic analysis would aim at identifying more systematically note-taker's ability and their difficulties to take their notes in French as a second language.

Key Words: Note-taking, criterion for coding, automatic data processing, L1/L2.

1. Introduction

L'objectif de ce travail est de présenter deux types d'approches méthodologiques pour le codage de corpus de prise de notes (désormais PDN) qui permettent, toutes deux, d'examiner les procédés utilisés par des noteurs natifs ou non-natifs du français.

La première approche s'inscrit dans une lignée classique de l'analyse des corpus de notes. Elle consiste en un codage manuel des données selon des descripteurs déterminés par les objectifs de la recherche en cours (cf. par exemple, Barbier, Faraco, Piolat, Roussey, & Kida, 2003, ce numéro). La nature des indices codés est décrite ci-après. Les interprétations possibles sur le fonctionnement des noteurs, à partir de ces indices, sont aussi explicitées.

La deuxième approche vise à coder les données des corpus de notes recueillis à l'aide du langage XML. Les données sont implémentées dans une base de données qu'il est possible d'interroger à l'aide d'une interface informatique (prototype PDN XML, Falaise, 2003 ; Tutin, Falaise, & Boch, 2003). Aussi, après un balisage manuel des données, cette interface permet d'en faciliter le traitement automatique selon les perspectives de recherche choisies. Les données ainsi analysées contiennent des annotations spécifiques à propos des différentes procédures de PDN (abréviations, substitution, spatialisation, etc.) qui ont été repérées dans des travaux antérieurs. Ce codage électronique devrait permettre la mise à disposition des données sous un format facilement exploitable et faciliter le repérage des procédés les plus fréquents.

Ces deux méthodes impliquent, en amont, un repérage et une description des indicateurs les plus généralement étudiés pour décrire les caractéristiques formelles des notes comme cela est réalisé pour la rédaction des textes (Piolat & Pélissier, 1998). Ce travail s'inspire donc des descriptions proposées dans la littérature, qu'il s'agisse d'études sur la configuration d'ensemble des notes (Faraco, Barbier, & Piolat, 2002 ; Piolat, 2001 ; Piolat, Roussey, & Barbier, 2003, ce numéro) ou sur les procédés locaux utilisés par les noteurs français (Boch, 1999 ; Branca-Rosoff, 1998, 2000, 2001 ; Faraco, 1997 a & b, 2000, 2003). La validité de ces indicateurs pour appréhender les stratégies et les performances de prises de notes est explicitée ci-après.

2. Approche traditionnelle : le codage des différentes variables

Ce type d'approche pour l'analyse des corpus de PDN s'inspire de la méthodologie telle qu'elle est traditionnellement pratiquée par les linguistes et psycholinguistes (Chaudron, Cook & Loschky, 1988 ; Dunkel, 1988 ; Branca-Rosoff, 1998 ; Piolat, 2001). Le travail d'analyse porte sur les corpus de notes originaux dont les caractéristiques de surface (par exemple, la quantité de mots notés, les types d'abréviations, etc.) sont choisies en fonction des objectifs de la recherche. Les données relatives à chaque corpus sont ensuite systématiquement catégorisées en fonction des variables retenues. Elles sont transcrites une à une sur une feuille de calcul de type Excel, de façon à être, par la suite, quantifiées et analysées.

Autrement dit, la démarche méthodologique consiste à déterminer dans un premier temps la nature des informations à étudier. Certaines informations sont analysées de façon récurrente en L1 comme en L2 (Boch, 1999, Branca-Rosoff 1998, sous presse ; Faraco, Barbier & Piolat, 2002a ; Faraco, Kida, Barbier & Piolat, 2002). Il s'agit, par exemple, des variables suivantes : les mots identiques au texte source (celui qui a été lu et entendu par les noteurs), les types d'abréviations (apocope, troncation, initialisme, etc.) ou encore les icônes observées dans les notes (logogramme, pictogramme). Pour chacune de ces variables, une fois définies, le travail de repérage et de codage des données est fait pour chaque corpus de notes.

Le tableau 1 présente une grille d'analyse correspondant à cette approche, telle qu'elle a été élaborée pour la recherche de Barbier et al. (2003, ce numéro). Dans une perspective de psycholinguistique, les variables choisies visent à rendre compte des procédés de sélection de l'information (ampleur et fidélité des prises de notes) et des procédés de retraitement de

l'information (procédés d'abréviation, de substitution et de structuration). Les choix qui ont présidé à la construction de cette grille et les façons de l'utiliser sont présentés dans la section 2 ci-après.

Tableau 1. Extrait de la grille d'analyse et exemples de codage des procédés de prise de notes issus de la recherche de Barbier, Faraco, Piolat, Roussey & Kida (2003, ce numéro)

	Mots notés				**Procédés de condensation**				**Procédés de structuration**	
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**
Mots du Texte source	**Mots Identiques**	**Mots Nouveaux**	**Mots en L1**	**Mots En L3**	**Abréviations Tronc. de la fin**	**Abréviations Charp. de consonne**	**Abréviations Complexes**	**Icônes substitutives**	**Marques de Liste**	**Marques de métadiscours**
Voici										
une										
brève										
information	Inf.				Inf.					
sur										
les	les									
procédures	proc.				Proc.					
d'accueil										
et									tiret	
d'inscription	inscription									souligné
des										
étudiants	étudiants									
étrangers	étrangers									
à										
l'Université	Université									
de		à								
Provence.		Marseille				Mrs				
(…)										
Total = 17	**7**	**2**	**0**		**2**	**1**	**0**	**0**	**1**	**1**
Proportions	**7/9**	**2/9**	**/**		**2/3**	**1/3**	**/**	**/**	**/**	**/**

Il faut noter qu'à l'issue du codage, avec ce type de méthode, les données obtenues concernent les seules variables définies en début d'étude. Si, ultérieurement, la modification d'une variable ou bien l'introduction d'autres variables s'avèrent nécessaires pour avancer dans la compréhension des procédés de PDN recueillies, chaque corpus doit alors être re-analysé.

2.1. Rendre compte du volume de notes produites et de leur fidélité

Dans la très grande majorité des travaux sur la PDN, un des objectifs est de repérer la quantité et la nature des informations que les noteurs ont sélectionnées pendant leur PDN. Il s'agit, dès lors, de comparer les caractéristiques des notes prises avec celles du texte source présenté aux noteurs (par oral ou par écrit). La nature de cette comparaison peut être de deux ordres et dépend du grain d'analyse choisi comparativement au texte source. En effet, l'analyse peut viser à cerner le stockage d'informations (grain large, en termes d'unités de sens) ou bien le stockage des mots (grain resserré autour des éléments de lexique).

2.1.1. Codage par unités lexicales

L'objectif est de recenser tous les mots que les noteurs ont introduit dans leurs notes. Un mot est considéré comme un groupe de lettres séparé de part et d'autre par un blanc, y compris les mots fonctionnels comme les articles, les pronoms, etc. ; cf. Chaudron, Cook, & Loschky, 1988)[1]. Cette définition du mot est celle qui est le plus souvent retenue dans la littérature (voir Barbier et al., 2003, ce numéro), à partir souvent de travaux portant sur les prises de notes en langue anglaise. Mais elle n'inclut pas l'apostrophe comme un caractère séparateur. Aussi, « l'université » correspond à un seul mot, alors que « les universités » en comprend deux. La spécificité de la langue française devrait donc conduire les chercheurs à définir le mot noté comme séparé par un blanc, ou par un apostrophe (sauf le mot aujourd'hui qui reste une exception).

En tous les cas, l'unité lexicale a été retenue comme étant pertinente pour les noteurs. Ils « écrivent des mots », ils définissent des mots de la même façon en L1 et en L2. Toutefois, cette conception remet en question les lexies complexes pour lesquelles les espaces ne sont pas valides (exemple, *au fur et à mesure*). La mise en ligne des mots du texte source dans la grille d'analyse peut cependant en tenir compte en regroupant ce type de lexies sur une même ligne.

La grille d'analyse présente en lignes successives tous les mots du texte source (colonne A, tableau 1). Chaque mot noté est consigné sur la même ligne que son correspondant dans le texte source (colonne B, tableau 1), qu'il s'agisse d'un mot abrégé ou non, et quelle que soit la langue de référence.

Une fois réalisé, ce codage permet d'élaborer des descripteurs dont la fonction est de quantifier l'activité. Plusieurs descripteurs peuvent être envisagés :

(a) *Le volume de notes produites* : Il peut s'agir des occurrences brutes de mots notés pour chaque noteur, ou bien de la proportion de mots notés par rapport au nombre de mots contenus dans le texte source. Ces descripteurs sont efficaces pour analyser la cadence de codage ou bien les procédures de sélection ;
(b) *La fidélité lexicale* : l'occurrence brute ou la proportion de mots notés (abrégés ou non) identiques à ceux du texte source ;
(c) *La fidélité syntaxique* : L'occurrence brute ou la proportion de mots notés fonctionnels (pronoms, articles, etc.) ou conceptuels (noms, verbes) peut présenter un certain intérêt pour distinguer des procédés de prise de notes linéaires (par exemple, « inscription des étudiants étrangers à l'université de Provence ») de procédés plus structurés autour des mots clés du texte source (par exemple, « inscription étrangers université » ; cf Annexe 1).

[1] « Criteria for the definition of a "word" in lecture notes follow Chaudon et al; (1988): that is all orthographic units with spaces on each side (except symbols) are treated as words. » (Clerehan, 1995, p. 140).

2.1.2. Codage par unités de sens

L'objectif est cette fois-ci de repérer toutes les unités de sens (éléments de contenu) que les noteurs introduisent dans leurs notes. Ce repérage implique au préalable une analyse de contenu du texte source. Parmi les procédés disponibles dans la littérature (Coirier, Gaonac'h & Passerault, 1996), le plus formalisé est celui développé par Kintsch et van Dijk (1988). Il s'agit de réaliser une analyse prédicative du texte source permettant d'en repérer les contenus propositionnels (en termes de Prédicat-Arguments) ainsi que les enchâssements rendant compte des relations hiérarchiques entre les propositions (c'est à dire la macrostructure du texte, comprenant les informations hiérarchiquement les plus importantes et leur expansions). L'application de cette technique est cependant longue et coûteuse, et il est possible de recourir à un autre procédé pour repérer les unités de sens et leur organisation hiérarchique : la méthode des juges (cf. à titre d'exemple, Barbier, Faraco & Piolat, 2002). Dans ce cas, il est demandé à plusieurs « juges » de délimiter intuitivement les différentes unités de contenu du texte source, ainsi que leur enchâssement. Les consignes qui leur sont données comprennent une définition de chaque type d'unité à repérer.

- *Les unités de sens de base* : une unité est censée apporter une seule information. Son format n'est pas conforme au format phrastique. Il peut varier d'un unique syntagme (fréquent) à un groupe de phrases (moins fréquent).
- *Les sous-unités de base* : une sous-unité de base comporte une information, mais elle complète une unité de base (qui la suit ou qui la précède). Il s'agit d'une expansion de l'information contenue dans une unité de base qui est de moindre importance. Le format de cette unité est variable (du syntagme à la phrase complexe).
- *Les unités conceptuelles* : Les unités de base et les sous-unités appartiennent à une unité englobante appelée unités conceptuelles. L'organisation syntaxique des énoncés et tout particulièrement les connecteurs permettent d'établir les frontières de ces unités. A chacune de ces unités il est possible d'associer un thème-titre (sorte de sous-titre) qui fait office de résumé.
- *Les unités majeures* : Les unités conceptuelles sont regroupées thématiquement. Elles peuvent être séparées explicitement ou non par des marqueurs syntaxiques (des connecteurs textuels comme *mais*, *donc*). Le contenu de ces unités peut, lui aussi, être spécifié au moyen d'un titre-résumé. Il s'agit des grands thèmes développés dans le texte.

Par la suite, soit les juges confrontent leur classement et établissent consensuellement l'organisation du texte source, soit les chercheurs repèrent, parmi les choix différents, les choix les plus fréquents qu'ils retiennent (pour une discussion sur cette méthode hiérarchique et les informations qu'elle ne prend pas en compte, cf. Boch, Tutin, & Grossmann, 2003, ce numéro).

Dans le cadre de la grille d'analyse présentée dans le tableau 1, les frontières des unités de sens peuvent être indiquées dans la colonne concernant les mots du texte source (colonne A). Il faut préciser que l'extrait présenté contient une liste de mots constituant une seule unité de base. A partir d'un tel codage, plusieurs descripteurs peuvent être envisagés :

(b) *Les thèmes notés avec précision* : après avoir fixé un critère d'ampleur des PDN (en nombre de mots par exemple), il est possible de calculer la proportion relative des unités de sens les plus abondamment notées comparativement à celles qui l'ont moins été.

(a) *La fidélité thématique* : Les occurrences brutes ou la proportion des unités de sens ayant fait l'objet d'une transcription (quel que soit le nombre de mots notés pour chaque unité). Les unités qui ont été jugées comme essentielles sur le plan thématique peuvent être analysées plus spécifiquement.

2.1.3. Portée interprétative de ces descripteurs

Les descripteurs par unités lexicales et par unités de sens sont complémentaires pour étudier les pratiques de PDN. Dans le cadre d'une étude comparative de la prise de notes en L1 et en L2, leur usage conjoint est crucial pour comprendre la nature des traitements et l'aisance avec laquelle les noteurs travaillent dans l'une et l'autre langue. Dans la situation de PDN en L1, noter un faible volume de mots ne traduit pas nécessairement une difficulté, mais plutôt une stratégie développée par le noteur, visant par exemple à retenir seulement les concepts les plus importants sous forme de mots-clés. Par contre, en L2, un moindre volume de notes peut indiquer que le noteur est en difficulté pour comprendre et stocker l'information. Une interprétation conjointe avec la fidélité syntaxique, lexicale et thématique permet de mieux comprendre cette difficulté. En effet, quand il éprouve des difficultés à noter en L2, le scripteur note les mots entendus sans introduire de mots nouveaux, et avec une syntaxe limitée. En outre, il peut percevoir plus ou moins bien l'importance des informations, et ne pas transcrire les unités qui ont été jugées comme étant essentielles dans le texte source.

Il faut noter que la fidélité renvoie à une stratégie délibérée dans le cadre de la PDN. Aussi, selon le contexte dans lequel les noteurs ont réalisé leur PDN, les différents indices de fidélité permettent de caractériser deux stratégies possibles. Le fait de noter le maximum d'éléments sans opérer de transformation immédiate est une stratégie classique des étudiants de licence observés par Boch (1999) en sciences du langage. Cette stratégie peut être focalisée sur des moments particuliers d'un enseignement, comme le mettent en évidence Parpette et Bouchard (2003, ce numéro) pour les définitions de termes juridiques. Cette stratégie de recherche d'exhaustivité peut même être culturelle comme le montre Omer (2003, ce numéro), les étudiants non natifs trouvant que les étudiants français notent abondamment. A l'inverse, la stratégie de fidélité peut être le signe d'une difficulté. Dans ce cas, les noteurs ne peuvent que « coller » à ce qu'ils entendent, sans opérer d'interprétation conceptuelle notamment. C'est, par exemple, le cas de noteurs japonophones qui notent en français L2 avec des difficultés de compréhension (Barbier, Faraco, Piolat, Roussey, & Kida, 2003, ce numéro).

2.2. Rendre compte des procédés de retraitement de l'information

L'analyse des prises de notes invite, par ailleurs, à élaborer des indicateurs sur la façon dont les noteurs ont compris et mis en forme l'information. En raison de l'écart important entre la cadence de parole rapide du conférencier et celle de saisie graphique lente du scripteur, et en raison des limites de stockage en mémoire de travail des informations entendues (Piolat, 2001 ; Piolat, Roussey, & Barbier, 2003, ce numéro), les noteurs utilisent pour noter plus rapidement des procédés de condensation. Il peut s'agir de procédés abréviatifs, qui sont de plusieurs types (cf. § 2.2.1. ci-après ; cf. Annexe 2). Ce peut être aussi des équivalents substitutifs, en changeant par exemple d'unité lexicale (logogrammes). Les noteurs utilisent aussi des marques de structuration ou de hiérarchisation des informations qu'ils transcrivent (pour un recensement des procédés abréviatifs, cf. Boch, 1999 ; Branca-Rosoff, 1998, sous presse). Par hypothèse, plus les noteurs ont des facilités de traitement (compréhension et mise en forme écrite), plus ils peuvent réaliser ce « re-traitement » de l'information pendant la PDN.

2.2.1. Descripteurs abréviatifs et substitutifs, et portée interprétative

Les procédés abréviatifs peuvent concerner chacun des mots notés, que ce soit un mot du texte source, un mot nouveau, voir un mot en langue étrangère (cf. Annexe 2). En outre, le chercheur peut créer autant de colonnes que de procédés qu'il veut identifier. Plusieurs indices ont

ainsi été utilisés pour coder les procédés de condensation en français langue seconde (voir tableau 1) :

- l'apocope (colonne F) ou troncature de la fin,
- la troncation des voyelles (colonne G) ou préservation de la charpente de consonnes,
- les procédés dits « complexe » (colonne H) pouvant combiner les deux premiers ou renvoyer à d'autres formats plus personnels comme la chute d'une syllabe et ou d'une lettre,
- l'utilisation de logogrammes (ou icônes substitutives d'unités lexicales, colonne I) qui sont des représentations symboliques de mots ou de locutions verbales, comme le point d'interrogation *?* pour *question* ; □ pour *psychologie* ; + pour *et*, etc.

Une fois réalisé, le codage de ces procédés permet d'opérer les calculs suivants :
(a) *La proportion de mots abrégés par rapport au nombre de mots notés*. Cette quantification permettra d'évaluer l'ampleur des pratiques abréviatives.
(b) *La proportion de chacun des procédés abréviatifs ou substitutifs*. Ces différentes quantifications renseignent sur les procédés de condensation les plus usités.

A ce stade des recherches, la valeur interprétative de ce type de descripteurs est hypothétique. Il ne s'agit pas pour le chercheur de décider qu'il est bien d'abréger beaucoup des mots notés. Mais l'usage d'abréviations et/ou de logogrammes relève d'une nécessaire rapidité d'exécution de la transcription. La disponibilité de ces procédés de condensation conditionnera donc la cadence d'écriture. Il s'agit là d'une expertise en termes de « transcodage ». Aussi, en raison des travaux sur l'importance des procédures automatisées en écriture (Kellogg, 1994), il est possible d'avancer que plus le noteur utilise de procédés abréviatifs, plus il note avec une certaine forme d'aisance. Cette orientation interprétative est compatible avec ce qui est observé chez les noteurs en français langue seconde : ils utilisent moins d'abréviations qu'en langue première (Barbier, Faraco, Piolat, Roussey, & Kida, 2003, ce numéro). Cependant, dans le cadre de la PDN en L2, ce type d'interprétation doit être nuancée : avant de considérer les difficultés dûes à la situation en langue étrangère, il faut tenir compte du fait que les pratiques de prises de notes peuvent être conditionnées par les caractéristiques structurelles de la langue écrite en L1. Par exemple, le procédé abréviatif par apocope (par exemple, *proc*. pour *procédure* ; *univ*. pour *université*, etc.) peut être employé pour noter dans tous les systèmes d'écriture alphabétique (espagnol, anglais, etc.), mais aussi dans les systèmes altaïques comme le japonais. Avec le codage en hiragana, le noteur japonais peut tronquer un ou plusieurs graphèmes syllabiques. Par contre, l'abréviation par préservation de la charpente de consonnes (par exemple, *dvpt* pour *développement* ; *cpdt* pour *cependant*, etc.) n'est adaptable qu'aux systèmes d'écriture alphabétique. Elle n'est pas possible en hiragana, système d'écriture syllabique où la distinction consonne-voyelle n'est pas valide.

Sur le plan descriptif, le codage systématique des différents procédés de condensation devrait permettre de repérer des régularités, des invariants. Les recherches peuvent être développées selon deux voies complémentaires. La première consiste à repérer la variété intra-sujet. Quel est l'éventail des procédures dont dispose chaque noteur ? Utilise-t-il de façon dominante certaines procédures plutôt que d'autres ? Les applique-t-il toujours aux mêmes unités lexicales ? La seconde voie consiste à repérer la variabilité inter-sujets qui devrait dépendre de facteurs endogènes (niveau d'habiletés des noteurs en L1 et/ou en L2, niveau de connaissances sur le contenu pris en notes, etc.), mais aussi de facteurs exogènes (style discursif du conférencier, etc.).

2.2.2. Pictogrammes et portée interprétative

Deux autres types de procédés peuvent enfin être codés pour rendre compte du retraitement des énoncés entendus (ou lus) par les noteurs (cf. Boch, 1999 pour une présentation organisée de ces procédés ainsi que Branca-Rosoff, sous presse ; cf. Annexe 1).

Les pictogrammes (colonne J, tableau 1) renvoient non pas à des unités lexicales mais à des relations logiques entre les unités sémantiques du texte. Ils sont associés généralement à des énoncés. Les pictogrammes les plus fréquemment employés sont des tirets, des flèches, ou des numérotations («1, 2, 3 ou A, B, C ») indiquant une structuration indentée et/ou chronologique des informations. Dans la grille d'analyse, les pictogrammes sont localisés sur la même ligne que le mot les précédant directement dans le corpus de notes (en colonne B).

Les corpus contiennent aussi des marques de métadiscours introduites par les noteurs (souvent sous forme de pictogramme, mais aussi de soulignement, encadrés, traits de séparation, renvois). Ces marques indiquent une prise de distance du noteur par rapport à ses notes (pour indiquer par exemple "c'est important" ou "ceci est à mettre en relation avec cela"). Elles sont codées dans la colonne K, en relation avec le mot qui précède directement cette marque dans le corpus de notes (en colonne B) et ce, même si le soulignement ou l'encadré implique plusieurs mots.

Une fois le codage établi, la quantification des données peut être réalisée à partir des simples occurrences observées car ces phénomènes sont généralement moins employés par les noteurs comparé aux procédés abréviatifs décrits ci-dessus.

Sur un plan interprétatif, les procédés pictographiques, principalement les effets de liste, permettent de statuer sur le niveau de compréhension des noteurs. Ces différents moyens de mise en forme matérielle des notes traduisent la capacité des noteurs à structurer et à mettre en relief des informations en se dégageant de la mise en forme linéaire imposée par la syntaxe.

2.2.3. Changement d'unité lexicale et portée interprétative

Coder, comparativement au texte source, les changements d'unité lexicale (cf. mots nouveaux, colonne C du tableau 1) peut être intéressant pour pister les retraitements d'information réalisés par les noteurs. Souvent, les noteurs transcrivent avec des mots qui leur sont propres un sens comparable (par exemple, "renouver" pour renouveler ou "change" pour échange). Ces interprétations se traduisent soit par une reformulation, soit par une création par rapport aux mots du texte original.

Dans la grille d'analyse, les mots nouveaux sont codés sur les mêmes lignes que les mots du texte source auxquels ils semblent correspondre sur le plan du contenu. Si le mot nouveau ne renvoie à aucun mot du texte source, il est codé sur la même ligne que le mot du corpus de notes le précédent (voir colonne B).

Une fois codé et quantifié, cet indicateur traduit la flexibilité des noteurs à comprendre et condenser de l'information. Il peut renseigner notamment sur leurs habiletés à développer une stratégie conceptuelle qui peut aller jusqu'à l'élaboration d'une carte conceptuelle (ou notation en mots-clés ; Slotte & Lonka, 2001 ; Piolat, 2001). En outre, cet indicateur semble particulièrement efficace pour contraster les stratégies de PDN d'étudiants hispanophones et japonophones en L1 et en français L2 (Barbier et al., 2003, ce numéro). Dans cette étude, les étudiants japonais ont employé en langue première une forte proportion de mots nouveaux, mais ils ont abandonné cette stratégie en L2. Les étudiants espagnols ont employé peu de mots nouveaux en L1 et encore moins en L2. Tout se passe comme si l'importance du changement lexical traduisait le degré de sûreté des étudiants. Pour les deux groupes d'étudiants, et particulièrement pour les étudiants japonophones ayant participé à l'expérience, cette sûreté est moindre en L2.

2.3. Rendre compte de l'alternance des codes linguistiques (L1, L3)

L'analyse des procédés de prises de notes par des non-natifs du français L2 requiert de prévoir l'apparition d'éléments linguistiques issus de leur première langue (et même parfois d'une langue tierce). A l'oral, de nombreux travaux attestent du fait que les locuteurs bilingues pratiquent fréquemment l'alternance de codes ou *code-switching*, en exprimant des parties plus ou moins importantes de leurs énoncés dans leur langue maternelle. Un phénomène similaire est aussi observé dans le cadre de la production écrite, les scripteurs en langue seconde pouvant faire référence à leur langue native durant leur production (Cumming, 1990 ; Manchon & al, 2000 : Wang & Wen, 2001). Il est donc possible que cet effet soit observé dans le cadre des activités de PDN, et ce particulièrement quand les noteurs sont en difficulté en langue cible.

Aussi, pour l'analyse des prises de notes en français L2, la quantité de mots notés en langue étrangère est un indicateur qui peut s'avérer pertinent. Les mots en langue étrangère sont de deux types : soit ils sont formulés en langue maternelle (L1, par exemple espagnol ou japonais), soit ils sont formulés dans une tierce langue quelle qu'elle soit (L3).

Sur le plan du codage, le mot identifié en L1 (colonne D du tableau 1) ou L3 (colonne E) est inscrit sur la même ligne que le mot du texte source (L2) auquel il semble correspondre sur le plan sémantique. Si le mot en langue étrangère ne correspond à aucun mot du texte source, il est situé sur la même ligne que le mot codé précédemment en colonne B.

3. Prototype PDN XML : pour un traitement automatique des corpus de notes

Une des limites importantes du codage manuel des corpus de prises de notes sur un tableau de type Excel (cf. section 2) réside dans la difficulté d'accès aux données originales une fois les corpus codés. Pour chaque PDN, le codage est réalisé sur autant de pages que l'impose le format du texte source, c'est-à-dire autant de lignes qu'il y a de mots. La récupération d'informations ponctuelles peut être faite via un affichage et une sélection de portions de tableau Excel. Elle est donc très fastidieuse. Excel permet l'élaboration de bases de données exploitables par procédures de tris dynamiques croisés afin de sélectionner les seules variables pertinentes. Mais ces procédures ne sont valides que pour des données numériques, alors que l'important serait de préserver la « visibilité » de ce que les noteurs ont écrits. Aussi, tout en utilisant les variables de codage présentées précédemment, le logiciel PDN XML a été mis au point (Falaise, 2003 ; Tutin, Falaise, & Boch, 2003) de façon à exploiter de façon plus souple ces données textuelles.

Pour le logiciel PDN XML, la procédure est la suivante (cf. Figure 1) :

(1) Les prises de notes sont balisées sous forme électronique à l'aide du langage de balisage XML qui permet de préciser les structures et marques spécifiques. Le balisage manuel est effectué à l'aide du logiciel XMetal.
(2) Les prises de notes en XML sont transformées à l'aide d'un script écrit en Perl de façon à être facilement exploitables par l'interface PDN XML.
(3) L'interface PDN XML a été réalisée entièrement en HTML et en JavaScript[2] pour exploiter les données générées. Le logiciel permet de présenter les prises de notes et d'obtenir les données des corpus dans des tableaux selon différents points de vue (comptage, sélections, etc.).

[2] L'inconvénient de ce système est qu'il se révèle assez lent pour l'utilisateur ; avec le recul, il semble qu'une implémentation en Java plus conventionnelle aurait été préférable.

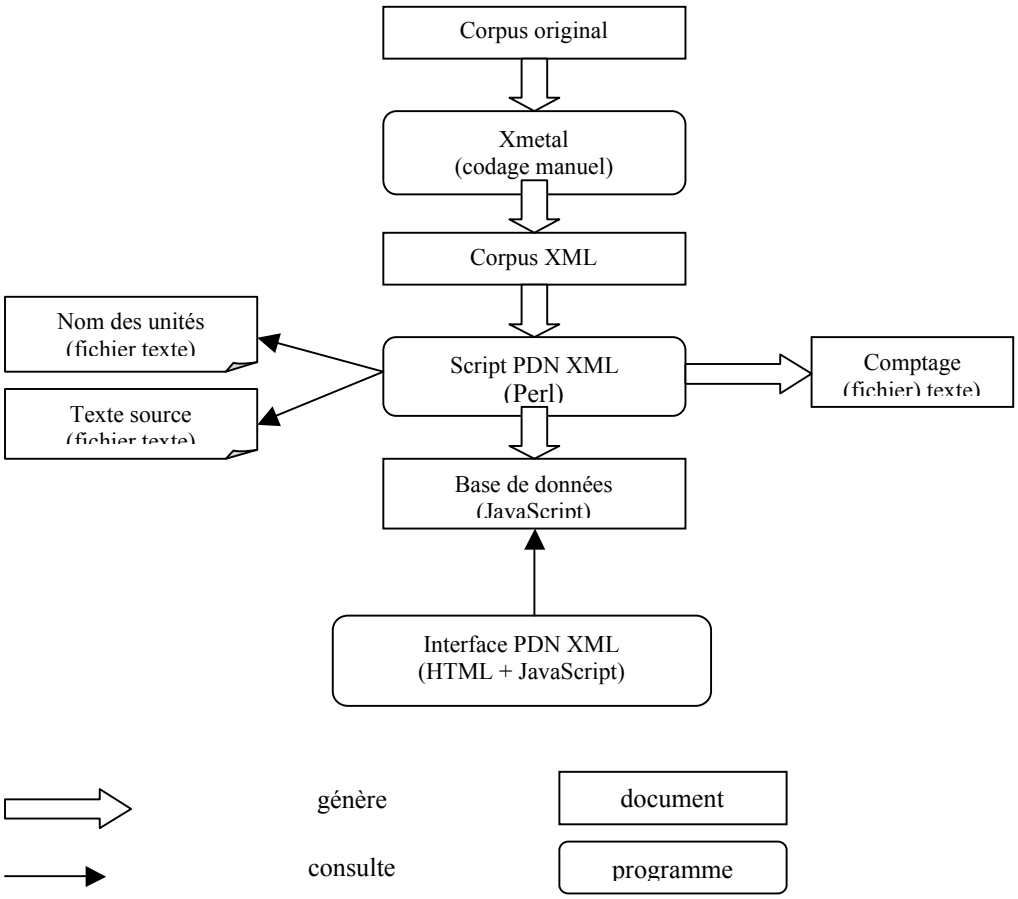


Figure 1 : Procédure d'utilisation du logiciel PDN XML

Pour fonctionner, l'interface nécessite un navigateur Internet. En théorie, n'importe quel navigateur peut permettre ce fonctionnement. En pratique, le bon fonctionnement de l'interface est garanti sous Internet Explorer 5+. De plus, les menus et la page principale utilisent des *cookies*[3] pour communiquer entre eux ; ceux-ci doivent donc être activés. En général, ils le sont par défaut dans tous les navigateurs.

3.1. Codage des données en XML

Pour pouvoir être traitées par le programme, les prises de notes ont été balisées en langage XML (eXtensible Markup Language), un langage de balisage normalisé élaboré par Word Wide Web consortium[4] très largement utilisé pour le traitement des corpus textuels et qui est appelé à remplacer HTML sur le Web.

[3] Fichiers qu'un serveur distant peut enregistrer sur un ordinateur extérieur. Les informations sont ainsi accessibles à un stockeur de pages web extérieur. Celui-ci pourra ainsi garder trace de toutes opérations qui auront été effectuées précédemment.

[4] www.w3.org/XML.

Le codage des données doit être fait à partir d'une structure préétablie. La structure retenue pour élaborer le prototype est celle issue de l'analyse du texte source (texte entendu par les preneurs de notes) exploité dans les recherches de Barbier, Faraco, Piolat, Roussey et Kida (2003, ce numéro) et de Boch, Tutin et Grossmann (2003, ce numéro). Avec la méthode des juges, ce texte a été catégorisé en unités de sens organisées hiérarchiquement (unités majeures, unités conceptuelles, unités de base). Le fichier correspondant au corpus d'un noteur est codé manuellement en XML, à l'aide du logiciel Xmetal, en suivant la structure du texte de base qui est à l'origine du texte source. Ainsi, chaque PDN a été balisée selon une DTD (définition de type de document) indiquant la correspondance avec le texte source et en particulier avec les principales unités. Les spécificités de la PDN ont, par ailleurs, été indiquées à l'aide d'éléments spécifiques pour les abréviations, les ratures, les fautes d'orthographe et termes en langue étrangère, les icônes, les imprécisions, les marques énumératives, les parties soulignées ou entourées. La figure 2 présente un exemple d'extrait de PDN annoté en XML.

```
<UMAJ>
        <UCONC>
        <UBASE><!--Si maintenant vous vous inscrivez par vos propres moyens, hors
          des  programmes d'échanges,--></UBASE>
        <UBASE><!--sachez que c'est possible dans tous les départements de
        l'université -->
       <ICONE>*</ICONE><ABREV>dns ts depmts</ABREV>
            <SUBASE><!--(que ce soit - les arts, - les lettres, - les langues,
            - les sciences humaines et sociales etc.).--></SUBASE></UBASE>
                <UBASE><!--Mais dans ce cas, l'inscription se fait de façon
individuelle,-->
           <ABREV>fçn indiv </ABREV>faut <ABREV>renouv</ABREV></UBASE>
        <UBASE><!-- et doit être renouvelée chaque année.--></UBASE>
        <UBASE><!--Le problème aussi, c'est que vous n'aurez pas d'aide pour les
aspects
                pratiques de la vie en France, -->
          pas d'aides <ABREV>asp </ABREV>vie <ABREV>quot</ABREV>
            <SUBASE><!—et de ce fait, de nombreux étudiants trouvent ce premier
contact avec
            l'université difficile. --></SUBASE><SUBASE><!--Ces étudiants disent
souvent (je
           les cite) --></SUBASE>
           <SUBASE><!--qu'ils ont perdu beaucoup d'énergie - à aller d'un bureau à
un autre
            avant d'avoir la bonne information, --></SUBASE>
           <SUBASE><!--- que personne n'était là pour les renseigner -->
           "perdre <ABREV>energ</ABREV>" <ICONE>- </ICONE><ABREV>pers. pr
            renseig.</ABREV></SUBASE>
           <SUBASE><!--- et qu'ils ont passé beaucoup de temps à comprendre ce
qu'il fallait
            faire, --></SUBASE>
           <SUBASE><!--- avant de vraiment pouvoir commencer leurs études.--
></SUBASE>
         </UBASE>
    </UCONC>
…
</UMAJ>
```

Figure 2. Un exemple de traitement de PDN en XML

Dans cet exemple, le texte source correspondant aux unités (majeures, conceptuelles, unités de base, sous-unités de base) apparaît dans des commentaires XML (en gras dans notre figure). Les extraits de PDN ont été insérés en fonction de leur position par rapport aux unités sémantiques du texte source. Les abréviations produites sont entourées de balises spécifiques

<ABREV> qui permettront des recherches ultérieures sur la base de donnée générée à partir des textes balisés.

3.2. Extraction des données

Dans un document XML, la recherche d'informations à l'aide d'une interface adéquate est relativement longue. Aussi, il était préférable d'effectuer un codage complet de chaque corpus, en générant une base de données contenant le plus possible d'informations. Cette étape de codage est réalisée par plusieurs programmes en langage Perl, langage bien adapté au traitement des corpus textuels.

En raison de cette saisie, ce fichier est alors susceptible de contenir des erreurs de structure. Aussi, un script en Perl a été prévu afin de vérifier si le fichier comporte les mêmes types d'unités que le texte source et dans le même ordre. En cas de divergence, par exemple si une unité a été « oubliée » dans un des fichiers, celui-ci devra être corrigé ou supprimé, sinon il ne pourra être traité.

Une fois la structure du fichier constituée, le programme principal lui attribue un nom à partir des informations données sur le scripteur et qui sont contenues dans l'en-tête. Dans la figure 3 et dans la première colonne est disponible l'identification des noteurs. L'avantage de ce système est de pouvoir ajouter un document nouveau. Placé dans le répertoire adéquat, ce fichier sera automatiquement pris en compte au prochain démarrage du programme et intégré aux autres fichiers.

Ensuite, le programme établit la structure et le nom des unités de découpage du corpus à partir d'un fichier texte (pour un exemple, cf. la première colonne de la figure 4). A chacune de ces unités correspond une partie du texte source qui, lui, est récupéré dans un autre fichier texte. L'interface peut donc en principe (cette possibilité n'a pas été testée) servir pour d'autres corpus de prises de notes associés à un autre texte source, s'ils ont été codés suivant les mêmes conventions.

Enfin, chaque document est analysé à l'aide d'expressions régulières visant à en extraire toutes les informations qui apparaîtront dans l'interface : le « texte brut » (c'est-à-dire l'intégralité du texte effectivement pris en notes), mais aussi les icônes, les abréviations, les imprécisions, les ratures, et les écarts. Les éléments appartenant à ces catégories spécifiques sont identifiés par des balises HTML pour permettre leur mise en valeur dans l'interface grâce à des CSS (Cascading Style Sheet ou feuille de style) qui évite de reformater chaque partie du texte. Ces informations sont alors écrites dans la base de données. Chaque type d'information est mis en valeur par une couleur particulière (par exemple le rouge pour les icônes ; cf. figure 3). Les couleurs sont définies dans une CSS externe aisément modifiable.

```
body {font-family: Verdana, Arial, Helvetica, sans-serif; font-size: 10pt}
table {font-family: Verdana, Arial, Helvetica, sans-serif; font-size: 10pt}
.subase {font-style: italic}
.comm {color: silver; font-size: 8pt}
.abrev {color: green}
.icone {color: red}
.ecart {color: blue}
.imprecis {color: purple}
.rature {text-decoration: line-through}
.scripteurinfo {color: white}
.scripteurinfotitre {color: white}
```

Figure 3. Feuille de style (CSS) indiquant le style (couleur, police de caractères, etc.) dans lequel l'information (icône, rature, etc.) sera formatée afin d'être visuellement très identifiable.

Au total, cette façon de procéder, à base d'expressions régulières, est assez rapide à exploiter. Mais toute mise à jour, c'est-à-dire toute modification des conventions de codage des prises de notes ou correction d'un bogue, est problématique[5].

3.3. Exploitation de la base de données à l'aide de l'interface PDN XML

Pour chaque fichier, un script spécifique a pour mission d'opérer des calculs sur les variables prédéfinies (nombre de mots du texte source, mots effectivement notés, icônes, abréviations, etc.). Il peut le faire pour les différents niveaux d'organisation du texte source. (unités majeures, unités conceptuelles, unités de base) permettant ainsi une analyse à plusieurs degrés de granularité. Le résultat de ces comptages est écrit dans un fichier texte exportable vers Excel pour des traitements statistiques plus poussés (cf. figure 4).

Script	*Unité*	*TSource*	*TBrut*	*Icone*	*Abrev*	*Imprec*	*Rature*	*Ecart*	
IMS	1_IA	20		7	1	0	0	0	0
FVP	1_IA	20		6	0	0	0	0	0
IFM	1_IA	20		6	0	0	0	0	0
RP	1_IA	20		10	1	4	0	0	0

Figure 4. Exemple de comptages obtenus et triés par unités (Script : code du scripteur ; Unité : nom de l'unité. TSource : nombre de mots dans le texte source ; TBrut : nombre de mots dans la prise de notes)

L'interface permet de sélectionner la partie de la base de données que le chercheur souhaite consulter. Par « partie », il faut entendre trois entités différentes, correspondant aux trois menus de l'interface :

- *Fichiers*

Ce menu permet de sélectionner et d'afficher le fichier de PDN à partir de l'identité de son scripteur. Tous les fichiers peuvent être sélectionnés en même temps (bouton « Tout »). Plus sélectivement, il est possible de ne consulter que des parties d'entre eux, par exemple, selon la langue.

La figure 5 montre l'interface quand le menu fichier a été ouvert (à gauche) et que les scripteurs anglais ont été sélectionnés. Dans la fenêtre principale (à droite), pour chaque scripteur, le nom de l'unité apparaît sur fond gris foncé, le texte source sur fond gris clair et la PDN sur fond blanc.

[5] Actuellement, la représentation du corpus, dans le programme, est faite sous forme de chaînes de caractères. Ces mises à jour pourraient être facilitées par une modélisation des données sous forme d'un arbre de balises XML représentées par des objets. L'élaboration d'une telle structure de données nécessiterait certes un travail de modélisation supplémentaire, mais devrait simplifier le traitement ultérieur du corpus.

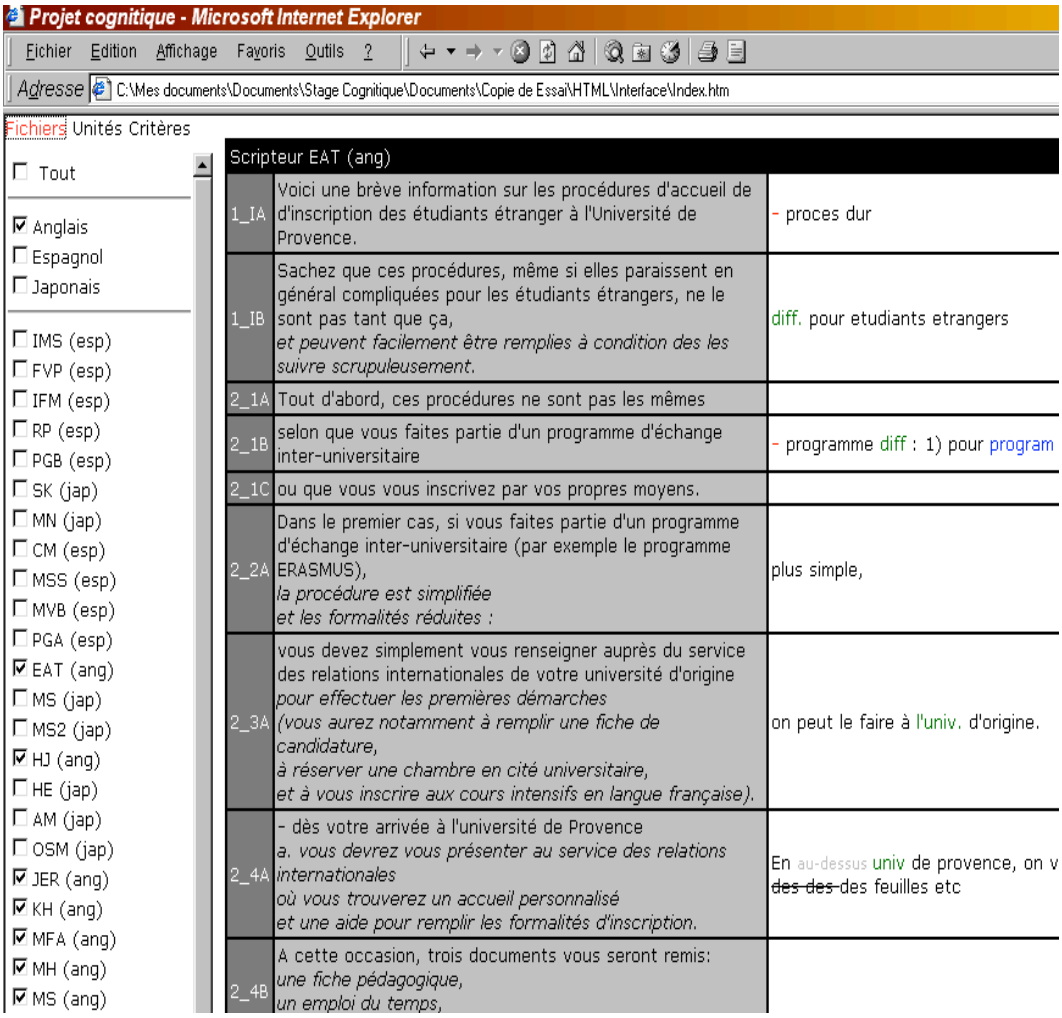


Figure 5. Exemple de vue de l'interface après la sélection d'un scripteur (EAT) dont la langue d'origine est l'anglais.

- *Unités*

Ce menu permet de sélectionner l'unité ou les unités (ou bien la ou les partie(s) du texte) que le chercheur souhaite consulter. Toutes les unités (bouton « TEXTE ») peuvent aussi être sélectionnées d'un coup

Cette dernière possibilité est illustrée dans la figure 6 où sont sélectionnés les mêmes fichiers que précédemment (cf. figure 5), mais cette fois c'est le menu unités qui est ouvert. La sélection a été limitée aux deux premières unités du corpus (cf. dans la colonne de gauche, 1_1A & 1_1B) qui apparaissent à droite pour les scripteurs anglais sélectionnés.

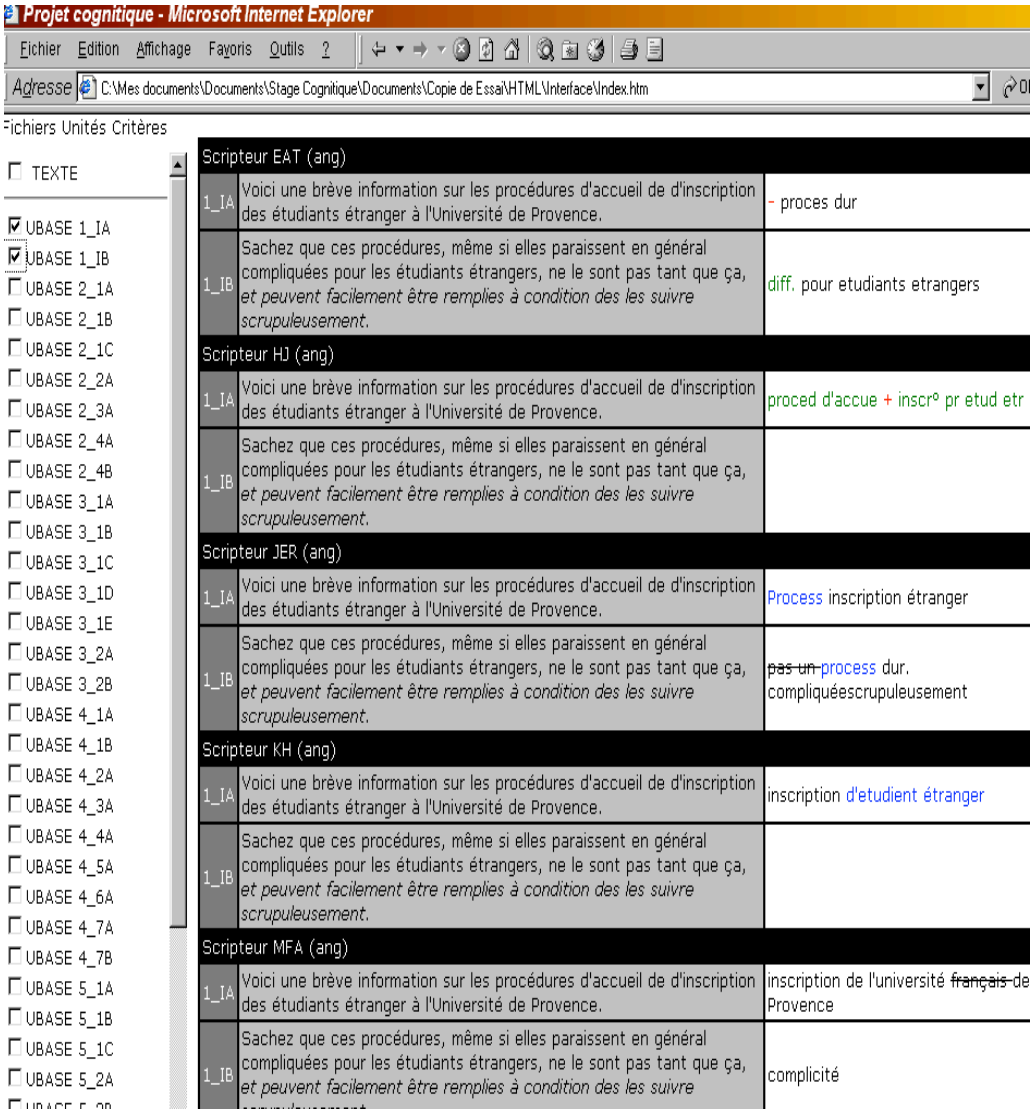


Figure 6. Vue de l'interface après la sélection des deux premières unités pour les scripteurs anglais.

- *Critères*

Ce menu permet de sélectionner le type d'information à afficher (Toute la PDN, seulement, les icônes ou les abréviations, etc.). Le bouton « Texte source » permet d'afficher ou non le texte source en regard des notes.

Dans la figure 7, toujours pour les scripteurs anglais, seules les abréviations sont affichées, sans le reste des notes et sans le texte source. L'affichage montre que seulement deux étudiants anglophones sur dix ont noté des abréviations dans leurs deux premières unités.

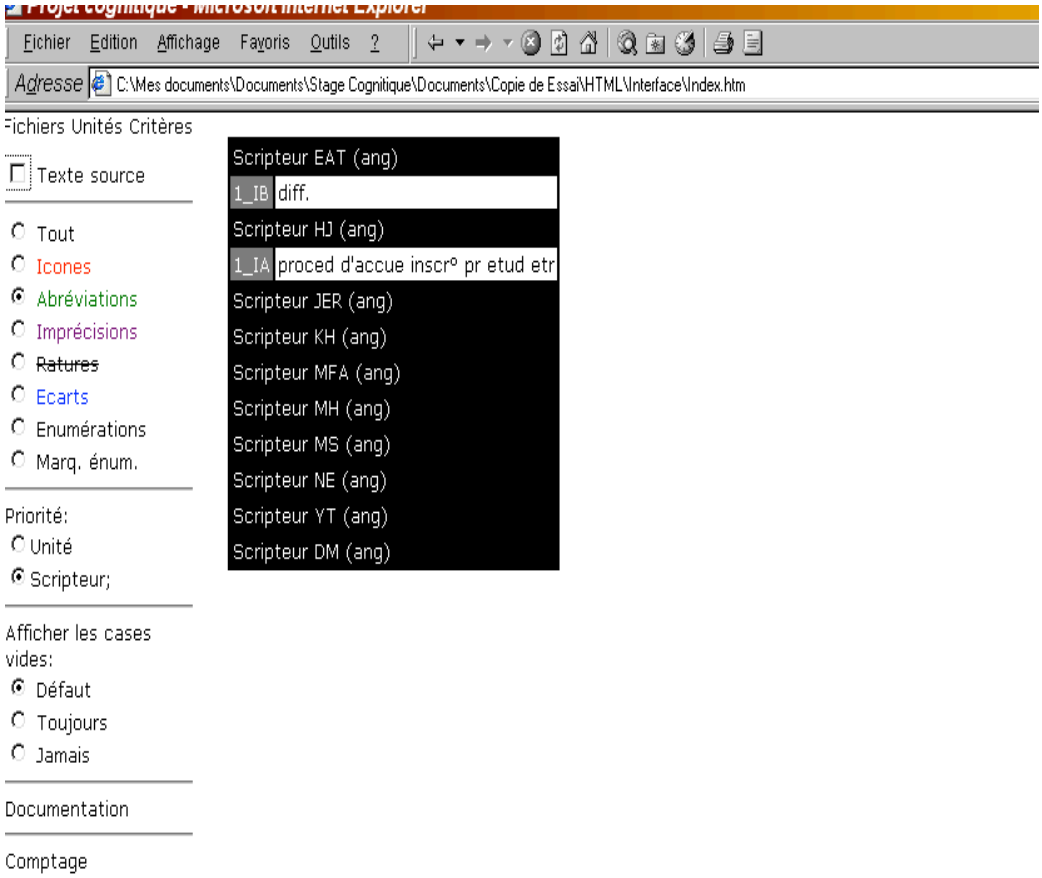


Figure 7. Vue de l'interface après la sélection du critère 'abréviations' pour les scripteurs anglais et pour les unités 1_1A et 1_1B.

L'option « Priorité » permet de choisir le type d'affichage par scripteur ou bien par unité. Dans la Figure 8, l'affichage est proche de celui de la figure 4, mais il est configuré à partir des unités.

Enfin, l'option « Afficher les cases vides » permet de masquer ou non les cases vides. De plus, grâce à des liens sur chaque nom d'unité, le chercheur peut afficher, au format JPEG, la prise de notes originales d'un scripteur.

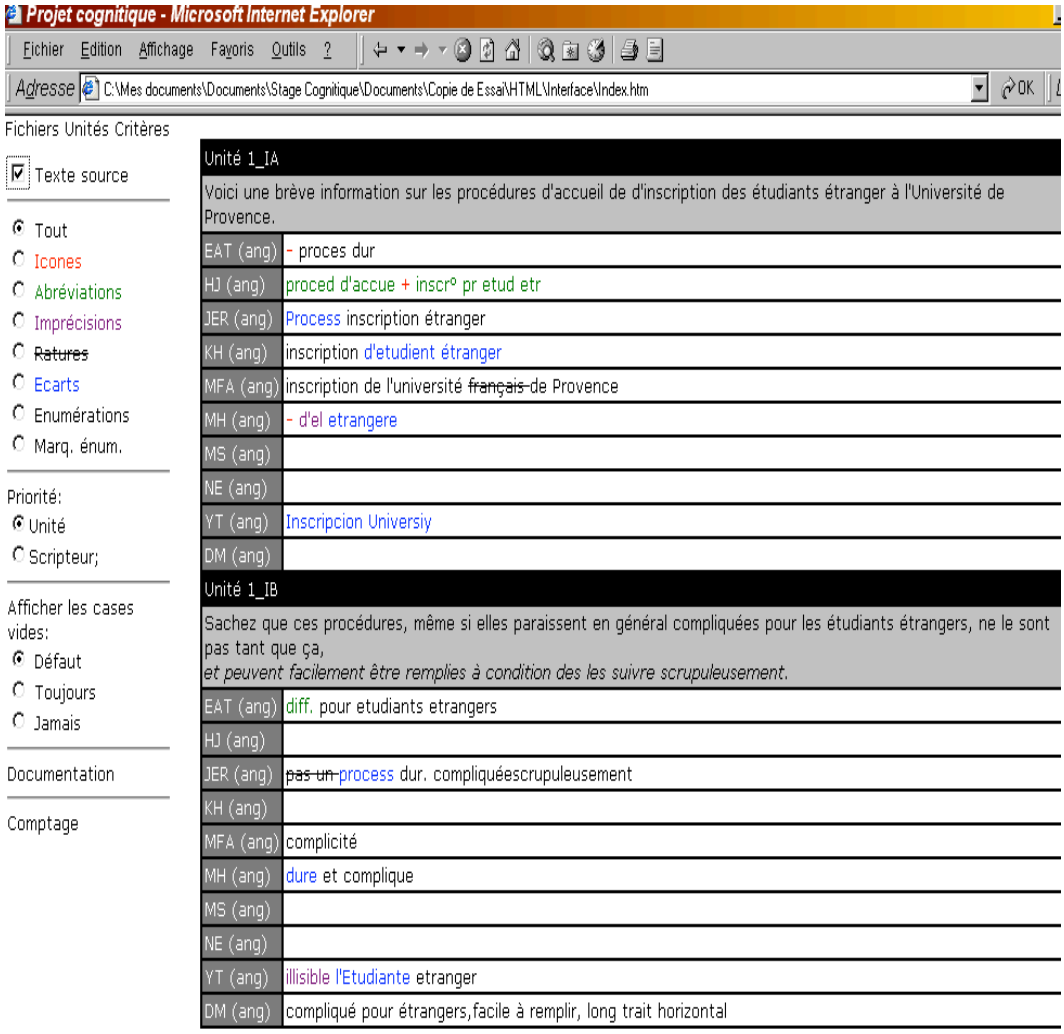


Figure 8. Vue de l'interface pour présenter unité par unité le critère 'abréviations' pour les scripteurs anglais.

4. Portées et limites de ces grilles d'analyse

Les codages proposés participent clairement à l'élaboration d'une grille d'analyse plus optimale pour effectuer le traitement des données des corpus de prise de notes. Il est possible dorénavant d'envisager l'analyse d'un grand échantillon d'informateurs, dont les PDN, après avoir été codées, pourront faire l'objet de différentes analyses. Le prototype PDN XML permet un important gain de temps pour réaliser des groupements et/ou des tris de données (cf. section 5.2.3.) qui simplifieront les analyses contrastives entre les groupes de noteurs.

Mais ce codage des corpus comporte plusieurs limites. Il est contraint par la structure en mots (retenue pour la grille Excel) et/ou par la structure en unités de sens (interface XML). Les données doivent être entrées manuellement et leur catégorisation impose au codeur une certaine part d'interprétation. Par exemple, ce dernier est amené à décider arbitrairement de la localisation dans la grille des mots ou des icônes qui ne correspondent pas au texte source, étant ainsi obligé

d'inférer à quel moment tel ou tel signe a été noté. De plus, pour les deux grilles d'analyse, il n'est pas toujours possible de rendre compte de l'organisation spatiale des notes, surtout quand leur distribution sur la feuille de papier n'est pas linéaire (avec par exemple des énoncés portés en transversal).

Par ailleurs, afin de poursuivre l'étude de la prise de notes en langue seconde (L2), la conception d'un mode de codage multilingue devrait être envisagée. Pour opérer un tel codage, le plus difficile est de trouver des critères qui s'appliquent à diverses langues, même quand celles-ci ne partagent pas le même système d'écriture. Les remarques qui vont suivre permettent de mesurer quelques-unes des difficultés envisagées pour la mise en place d'une grille d'analyse multilingue.

Barbier et al. (2003, ce numéro) ont comparé les notes prises par des étudiants hispanophones et japonophones en L1 et L2. Dans le cadre de cette étude, le découpage du corpus de notes produits en japonais (L1) a été difficilement ajustable au découpage effectué sur les corpus en français. En effet, même si la segmentation en mots se fait en japonais comme en français, la définition de ce qui constitue une unité lexicale (cf. 2.1.1.) doit être reconsidérée pour le japonais. Par exemple, il n'y a pas dans cette langue d'espace entre un mot et la particule fonctionnelle qui l'accompagne. En effet, en japonais, une phrase se compose d'un certain nombre de graphèmes, appartenant à trois systèmes d'écriture (hiragana, katakana, kanji). L'unité lexicale *université* s'écrit ainsi en japonais avec deux caractères en kanji "GRAND" (大, 3 traits) et "ÉTUDE" (学, 8 traits). Par exemple, l'adjectif *grand*, n'a de véritable statut lexical en japonais que lorsqu'il est combiné avec un suffixe. Tout se passe comme s'il s'agissait, en français, d'une forme comme "grand-" à laquelle il faudrait ajouter des particules comme "-ir" (grandir) ou "-eur" (grandeur) pour que cette forme ait un statut lexical. Ainsi, le caractère "GRAND" doit être combiné avec deux caractères en hiragana (き, 4 traits et い, 2 traits) pour former l'adjectif *grand* (大きい, infinitif[6])

Par conséquent, la segmentation lexicale du texte source en japonais nécessite une connaissance globale du fonctionnement des caractères, à la fois sur l'axe syntagmatique et sur l'axe paradigmatique. Plus précisément, la question des mots fonctionnels, qui peuvent être pertinents à analyser dans les prises de notes en français L1, doit être réfléchie pour le japonais. En effet, les Français utilisent des graphies fonctionnelles monosyllabiques comme *à* (préposition), *y* (pronom locatif), *le*, *la*, *les* (articles) alors que les Japonais ne disposent pas d'articles définis et indéfinis et que leurs déterminants sont limités à des adjectifs démonstratifs ou indéfinis. En outre, un certain nombre de mots fonctionnels (ou « particules grammaticales ») s'écrivent avec un caractère comme les particules monosyllabiques courantes : il y a *ga* (ergatif), *ha* (nominatif) *no* (datif), *mo* (nominatif-additif), *wo* (acusatif) *he* (locatif-direction), *ni* (locatif), *de* (instrumental), *te* (coordination). Le problème est que ces particules sont intégrées au mot et qu'il ne se dessine donc pas de correspondance sur le plan des mots fonctionnels entre le français et le japonais[7].

Pour résumé, l'essentiel des difficultés de codage en japonais : certains caractères purement fonctionnels n'indiquent qu'une fonction grammaticale à l'intérieur de la phrase ; d'autres renvoient à une entité en référence à un objet concret ou un concept abstrait ; d'autres encore nécessitent d'être combinés avec d'autres mots pour être reconnus au titre d'unité conceptuelle autonome. Par conséquent, si le système de segmentation du corpus en unités lexicales semble

[6] Pour plus de précisions, l'adjectif japonais est une catégorie grammaticale qui subit une flexion selon le type de catégorie grammaticale qui le suit. Par exemple, "pas grand" est "ooki-ku nai" (la forme suspensive), "grand+passé" "ooki-ka'-ta" (la forme post-prédicative) "devenir grand" ooki-ku naru, "grand. (infinitif) " est "ooki-i" (la forme conclusive), "homme grand" est "ooki-i hiko" (la forme post-nominale), "si grand", ooki-kere-ba" (la forme conditionnelle).

[7] Ceci est à rapprocher de l'intégration effectuée dans la recherche de barbier et al. (2003, ce numéro) entre le déterminant « l' » et le mot qu'il accompagne (cf. tableau 1, « l'université »). En japonais, cette intégration est sans doute plus fréquente.

adapté au traitement de langues comme le français, il peut au contraire neutraliser les spécificités grammaticales d'autres langues comme le japonais. La solution serait peut-être de définir comme base une unité syntaxique en français pouvant être constituée de plusieurs mots (la bibliothèque, l'étude des signes, tout d'abord, …). Dans ce cas, la particule japonaise pourrait être considérée comme une partie d'unité conceptuelle et une meilleure correspondance entre unités conceptuelles d'une langue à l'autre pourrait alors être trouvée. Il faut préciser que le découpage en unités conceptuelles de ce type reste encore à définir. En outre, il demanderait un repérage manuel dans les prises de notes avant la mise en grilles et nécessiterait sans doute l'utilisation de la méthode des juges.

5. Conclusion

Les descriptions qui viennent d'être faites sur le codage de corpus de PDN et sur les interprétations associées aux descripteurs génèrent plusieurs constats qui pourront permettre de poursuivre la réflexion afin d'améliorer ce type d'outil. Le dépouillement préliminaire des corpus de notes ne peut se faire sans choix de la part du codeur (unités, descripteurs, etc.). Ces choix dépendent des objectifs de recherche.

En explicitant les variables de codage les plus pertinentes, ce travail visait à repérer des invariants dans les pratiques de prises de notes en L1 et en L2. Le logiciel PDN XML mis au point par Tutin, Falaise, & Boch (2003) contribue à cet objectif. Mais la réflexion doit être poursuivie et d'autres variables restent à déterminer, pour conduire au mieux cette recherche d'invariants. Cette réflexion est cruciale, si l'objectif est de se donner les moyens de comparer des corpus de prise de notes de scripteurs non natifs en français L2 et dans leur langue d'origine.

Références

Barbier, M.-L., Faraco, M., Piolat, A., Roussey, J.-Y., & Kida, T. (2003).Comparaison de la prise de notes d'étudiants japonais et espagnol dans leur langue native et en français L2. *Arob@se 7*, 1-2 [http://www.arobase.to/v7/].

Boch, F. (1999). *Pratiques d'écriture et de réécriture à l'université - La prise de notes, entre texte source et texte cible.* Lille : Presses Universitaires du Septentrion.

Boch, F., Tutin, A., & Grossmann, F. (2003). Analyse de textes réécrits à partir de prise de notes. Intérêts de la méthode RST (Rhetorical Structure Theory). *Arob@se 7*, 1-2 [http://www.arobase.to/v7/].

Branca-Rosoff, S. (1998). Abréviations et icônes dans les prises de notes des étudiants. In M. Bilgerr, K. Eynde, & F. Gadet. (Eds.), *Analyse linguistique et approches de l'oral Recueil d'études offert en hommage à Claire Blanche-Benveniste* (pp.288-299). Paris : Peeters Leuven.

Branca-Rosoff, S. (sous presse). Les abréviations lexicales dans les prises de notes des étudiants. *Faits de langue.*

Chaudron, C., Cook, J., & Loschky, J. C. (1988). *Quality of lecture notes and second language listening comprehension*. (Technical Report, n°7). Honolulu: University of Hawaii, Center for Second Language Classroom Research, Social Science Research Institute.

Clerehan, R. (1995). Taking it down. Notetaking practices of L1 and L2 students. *English for specific purposes, 14 (2)*, 137-155.

Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written communication, 7(4),* 482-5111.

Dunkel, P. (1988). The content of L1 and L2 students lecture notes and its relation to test performance. *TESOL Quaterly, 22(2)*, 259-281.

Falaise, A. (2003). *Un logiciel de codage et de traitement automatique des corpus de prise de notes : le prototype PDN XML.* Grenoble : LIDILEM, EA 609, Université Stendhal.

Faraco, M. (1997a). Étude longitudinale de la prise de notes d'un cours universitaire français : le cas d'étudiants étrangers d'un cursus européen. *ASp, 15-18*, 41-54.

Faraco, M. (1997b). Technique de prise de notes en français spécialisé. *Le Français dans le Monde, 287*, 38-40.

Faraco, M. (2000). Prise de notes : quelles compétences pour les Européens ? In L. Collès, J.-L. Duffays, G. Fabry, & C. Maeder (Eds.), *Didactique des langues romanes. Le développement de compétences chez l'apprenant* (pp. 107-112). Bruxelles: De Boeck, Duculot.

Faraco, M. (2003). Analyse des paramètres de l'activité de prise de notes en langue seconde : une étude pilote. *Travaux Interdisciplinaires de Parole et Langage d'Aix-en-Provence (TIPA), 21*, 45-62

Faraco, M., Barbier, M. L., & Piolat, A. (2002). A comparison between L1 and L2 note-taking in undergraduate students. In G. Rijlaarsdam (Series Ed.), *Studies in Writing,* & S. Ransdell, & M. L. Barbier (Volume, Eds.), *New Directions for Research in L2 Writing* (pp. 145-167). Dordrecht: Kluwer Academic Publishers.

Faraco, M., Kida, T., Barbier, M.-L., & Piolat, A. (2002). Didactic prosody and notetaking in L1 and L2. In B. Bel, & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 conference* (pp. 287-290). Aix-en-Provence: Laboratoire Parole et Langage, Université de Provence.

Kellogg, R. T. (1994). *The Psychology of Writing.* New York: Oxford University Press.

Manchón, R.M., Roca de Larios, J. & Murphy, L. (2000). An approximation to the study of backtracking in L2 writing. *Learning and Instruction, 10,* 13-35.

Omer, D. (2003). La prise de notes à la française pour des noteurs non natifs. *Arob@se 7*, 1-2 [http://www.arobase.to/v7/].

Piolat, A. (2001). *La prise de notes*. Paris : PUF.

Piolat, A., & Pélissier, A. (1998). Etude de la rédaction de textes: contraintes théoriques et méthodes de recherches. In A. Piolat & A. Pélissier (Eds.), *La rédaction de textes. Approche cognitive* (pp. 225-269). Lausanne: Delachaux & Niestlé.

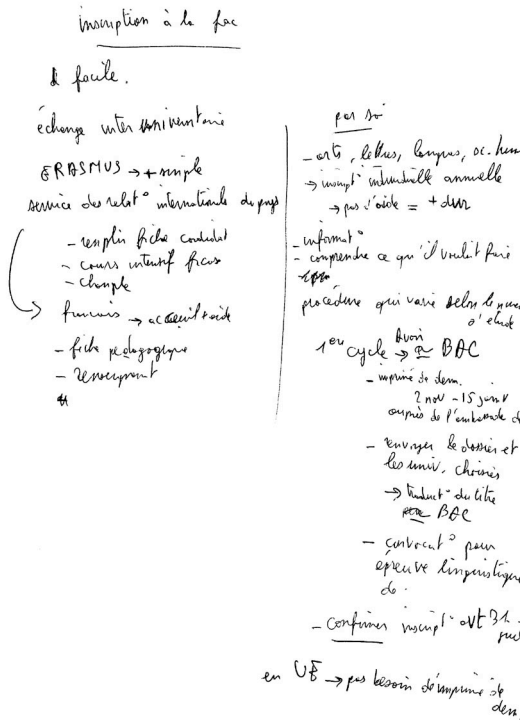
Piolat, A., Roussey, J.-Y., & Barbier, M.-L. (2003). Mesure de l'effort cognitif : Pourquoi est-il opportun de comparer la prise de notes à la rédaction, l'apprentissage et la lecture de divers documents ? *Arob@se 7*, 1-2 [http://www.arobase.to/v7/].

Slotte, V., & Lonka, K (2001). Note-taking and essay writing. In G. Rijlaarsdam (Series Ed.) & P. Tynjälä, L. Mason & K. Lonka (volume Eds.*), Studies in Writing, vol. 7, Writing as a learning tool: Integrating Theory and Practice* (pp. 131-141).). Dordrecht: Kluwer Academic Publishers.

Tutin, A., Falaise, A., & Boch, F. (2003). *Traitement informatique des corpus de prise de notes et des textes cibles.* Grenoble : LIDILEM, EA 609, Université Stendhal.

Wang, W., & Wen, Q. (2001). L1 use in the L2 composing process: an exploratory study of 16 chinese EFL writers. *Journal of second language writing, 11*, 225-246.

Annexe 1. Exemple de mise en forme matérielle des notes faites par un étudiant natif en L1.



Annexe 2. Abréviations en français (L1) et en anglais (L2) faites par le même étudiant.