

Lexical verbs in academic discourse: a corpus-driven study of learner use

Sylviane Granger & Magali Paquot

[DRAFT]

1. Introduction

In spite of their relative infrequency in English for Academic Purposes (EAP) as compared to other genres, notably conversation and fiction (Biber et al 1999: 358), lexical verbs contribute significantly to some major EAP functions such as expressing personal stance, reviewing the literature, quoting, expressing cause and effects, summarizing and contrasting. They enable writers to modulate their ideas and position their work in relation to other members of the discipline. Hinkel (2004) classifies them into the following five categories: activity verbs (*make, use, give*), reporting verbs (*suggest, discuss, argue, propose*), mental/emotive verbs (*know, think, see*), linking verbs (*appear, become, keep, prove*) and logico-semantic relationship verbs (*contrast, follow, cause, illustrate*). Among those it is undeniably the category of ‘reporting verbs’ that has received the most attention (Thompson & Yiyun 1991, Shaw 1992, Thomas & Hawes 1994, Hyland 1999, Charles 2006a & 2006b). Reporting verbs are important in academic discourse, as ‘they allow the writer to clearly convey the kind of activity reported and to precisely distinguish an attitude to that information, signaling whether the claims are to be taken as accepted or not’ (Hyland 1999: 344). Other categories, such as that of ‘coming-to-know verbs’ (Meyer 1997; Hiltunen 2006), have also been the subject of detailed investigation. In general, EAP studies have tended to focus on one specific category of verbs rather than give a general overview of the use of lexical verbs in academic discourse. Williams (1996) is an exception in this respect as he investigates all lexical verbs of a particular frequency used in medical reports.

Although verbs figure prominently in EAP word lists, it is difficult to draw up a list of EAP lexical verbs from existing lists as they often fail to give any indication of word category membership. The most popular EAP list, Coxhead's (2000) Academic Word List (AWL), contains many words like *conduct*, *focus*, *approach*, *survey* or *function*, which can be nouns and verbs. Another characteristic of the AWL is that it excludes the top 2,000 words in the language, i.e. those that figure in the General Service List (GSL). This is justified to some extent as many high frequency verbs are rarely used in EAP. For example, Biber (1988: 105) demonstrates that 'private verbs' like *love*, *want*, *like*, *feel* or *hope*, which 'are used for the overt expression of private attitudes, thoughts, and emotions', are typical of involved discourse, notably conversation, and rarely used in academic texts. However, there are in fact several high frequency verbs which turn out to play a major role in EAP and are therefore worth including in EAP syllabuses. For example, Meyer's (1997) study of the acquisition of knowledge in the process of academic investigation, includes high frequency verbs like *find* or *show* which show 'all the vaguenesses, polysemies, and ambiguities of everyday language', but 'are used to discuss matters lying at the very heart of the scholarly process' (ibid: 368). In order to give these verbs the coverage they deserve in EAP, Paquot (2007) has included in her Academic Keyword List (AKL) verbs like *aim*, *argue*, *cause*, *claim*, *effect* or *suggest*, which are absent from Coxhead's list.

Insufficient knowledge of EAP verbs is a serious handicap for learners as it prevents them from expressing their thoughts in all their nuances and couching them in the expected style. As pointed out by Swales (2004: 17), 'a formal research report

written in informal English may be considered too simplistic even if the actual ideas and/or data are complex'. Presenting learners with lists of EAP verbs and the exact meanings they convey is therefore undoubtedly an important first step but unless it is complemented with a detailed description of their use, results are bound to be highly disappointing. One of the strengths of EAP verbs, their ability to help modulate the message via tense, aspect, mood and voice, creates a minefield of difficulties for learners (Hinkel 2002, Swales and Feak 2004). Research has tended to focus largely on these areas of difficulty, in particular on the issue of tense and aspect and the question of the transferability of General English rules to EAP (Swales 1990: 151). However, this is not the only problem that learners are faced with. They also have to deal with the fact that each EAP verb has its own preferred lexico-grammatical company, viz subjects (*this study shows that; the evidence suggests that; these results suggest that,*), objects (*SUPPORT the view / hypothesis that ..., PROVIDE evidence / information*) and adverbs (*DIFFER significantly; VARY considerably / widely; APPLY equally; closely related; widely used; generally accepted*) and tend to appear in routinized structures (*as discussed in; there is (no, some, little) evidence that, it should be noted that*). Generalities such as 'The passive is very frequent in academic discourse' are not very helpful as some EAP verbs are hardly ever used in the passive while others are typically (if not exclusively) used in the passive (cf. Swales 2004: 12).

Lexico-grammatical restrictions of EAP verbs are often disregarded in EAP textbooks, which tend to present verbs separately from nouns and adverbs when in fact, as demonstrated by several recent learner corpus-based studies, it is their interaction that causes difficulty for learners. This is confirmed by Nesselhauf's

(2005) investigation of German-speaking English as a Foreign Language (EFL) learners' misuse of collocations in verb-noun combinations. Similarly, Hyland's (2008) analysis of word clusters in Cantonese-speaking students academic writing shows that 'many of the clusters most frequently used in published academic writing were never, or only rarely, found in the student texts' (see also Altenberg and Granger 2001, Ädel 2006).

All these studies show that it is phraseology in the wide sense, viz including both highly fixed and much looser routinized sequences, that EFL learners find most difficult. Some of these phraseological difficulties, in particular those related to pragmatic appropriacy and discourse patterns, are shared by novice native writers. Hyland and Milton (1997: 192) show that both Cantonese learners and novice native writers mix 'informal spoken and formal written forms and transfer conversational uses of academic genres'. Similarly, Neff et al (2004: 152) compare the expression of writer stance in various corpora of argumentative texts written by EFL learners, novice and professional native writers and show that 'all of the student writers (native and non-native) have the novice-writer characteristic of excessive visibility'. However, it would be wrong to conclude that native student writers and English as a Foreign Language (EFL)/English as a Second Language (ESL) learners face exactly the same difficulties in academic writing and can therefore be considered as belonging to one and the same category of novice writers. As pointed out by Gilquin et al (2007a) and further argued below, a wide range of lexico-grammatical difficulties are exclusive to L2 learners and therefore deserve specific attention.

The main objective of this chapter is to give a detailed description of the use of lexical verbs in L2 learners' academic writing compared to both expert and novice native writing. The investigation is based on native and learner corpora of academic writing and the method is corpus-driven rather than corpus-based, i.e. 'relies heavily on data and (largely) automatic procedures' (De Cock 2003:197) (cf. also Tognini-Bonelli 2001). The investigation attempts to tackle the following questions: Which (categories of) verbs do learners use in their EAP writing? Is the set of EAP verbs used by L2 learners different from both expert and novice native users? Do L2 writers use EAP verbs in their typical lexico-grammatical patterning?

In section 2 we describe the corpora and the methodology used to extract EAP verbs. Section 3 discusses the advantages and disadvantages of taking word forms or lemmas as units of analysis. Section 4 gives the results of the analysis of lexical verbs in EFL and professional academic writing. Section 5 addresses the issue of text type and domain comparability by revisiting the findings of section 4 in the light of a comparison between EFL and native novice writing. Section 6 contains concluding remarks.

2. Data and methodology

This study makes use of two large collections of academic discourse to describe the use of EAP verbs by native and learner writers. The learner data comes from the second edition of the *International Corpus of Learner English* (henceforth ICLE) (Granger et al. forthcoming) which contains over 3 million words of argumentative essay writing by high-intermediate to advanced EFL university students of 16 different mother tongue backgrounds: Bulgarian, Chinese, Czech, Dutch, Finnish,

French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Tswana and Turkish. The focus of our study is on EFL learners rather than ESL students. The two populations are rarely distinguished in the literature and yet they are quite different. For example, the use of phrasal verbs instead of the more EAP-appropriate single word equivalents is often presented as a major problem for EAP students (cf. e.g. Swales & Feak 2004). It may well be a problem for ESL learners exposed to informal English on a daily basis or for novice native writers who may transfer their everyday English to their academic texts. However, it is not a major source of difficulty for EFL learners, who make scant use of phrasal verbs (cf. Sjöholm 1998; Liao and Fukuya 2004).

A large collection of expert writing, which will be referred to as ACAD, is used as a comparable corpus. It is composed of the academic sub-parts of the MicroConcord corpus collection (Johns & Scott 1993) and the Baby British National Corpus (cf. Burnard 2003), which combined, contain 2 million words. Both corpora consist of published academic prose (book samples and articles) and are divided into five sub-corpora of c. 200,000 words, each of which corresponds to a broad academic discipline (e.g. humanities, social science, applied science, technology and engineering).

The main advantage of these two corpora is that they are large collections of academic texts and thus highly valuable in providing a general overview of the use of lexical verbs in academic writing. An important caveat however, is that the two corpora are not fully comparable. Expert texts are expository in nature, i.e. they are topic-oriented (cf. Britton 1994) and rely on the comprehension of general concepts

(cf. Werlich 1976) while argumentative essays ‘depart from the assumption that the receiver’s belief must be changed’ (Gramley and Pätzold 1992:193). In addition, expert texts are discipline-specific while learners’ essays discuss a range of general topics such as feminism, the impact of television, drugs, etc. Special care therefore needs to be taken to interpret results in the light of genre analysis as some differences between learner essays and expert texts may simply reflect differences in their communicative goals and settings (cf. Neff et al. 2004). Another issue concerns the use of professional native writing as a standard of comparison in learner corpus research. This has been criticized by several authors, among others Lorenz (1999: 14), who considers this practice to be ‘both unfair and descriptively inadequate’ and Hyland and Milton (1997: 184) who take a stand against the ‘unrealistic standard of “expert writer” models’ and argue that native student writing is a better type of comparable data to EFL learner writing if the objective is to describe and evaluate interlanguage(s) as fairly as possible¹.

To address these issues of comparability, two additional corpora are used in the second stage of our investigation. They have the advantage of representing the same text type, namely argumentative essay writing, and contain data from EFL learners and native novice writers. The learner corpus is a subcomponent of *ICLE* that only contains data from French-speaking learners. The corpus of student writing is a subpart of the *Louvain Corpus of Native Speaker Essays* (LOCNESS) (cf. Granger 1996), which consists of argumentative essays written by American university students. The two corpora are approximately the same size (150,000 words) and cover similar topics (e.g. *Crime does not pay*, *Feminists have done more harm to the cause of women than good*, *Most university degrees are theoretical and do not prepare*

students for the real world, and *In the words of the old song, money is the root of evil*). Table 1 gives an overview of the four corpora used.

Corpora	Number of words	Professional status	L1 or L2	Text type
ACAD	2,027,880	Professional	L1 / Proficient L2 writer	Expository
ICLE	3,233,214	Non-professional	L2 learner	Argumentative
LOCNESS	150,166	Non-professional	L1 novice writer	Argumentative
ICLE-FR	160,530	Non-professional	L2 learner	Argumentative

Table 1: Description of the corpora

All corpora were lemmatized and part-of-speech tagged with the *Constituent Likelihood Automatic Word-tagging System* (CLAWS) C7 (cf. Garside and Smith 1997).² Figure 1 shows an example of CLAWS C7 horizontal output: each word form is followed by its part-of-speech (POS) tag. The tagset includes six different tags for lexical verbs: VV0 (base form, e.g. *drink*, *work*), VVD (past tense, e.g. *drank*, *worked*), VVG (-ing participle, e.g. *drinking*, *working*), VVI (infinitive, e.g. *drink*, *work*), VVN (past participle, e.g. *drunk*, *worked*), VVZ (-s form, e.g. *drinks*, *works*).

The_AT whole_JJ point_NN1 of_IO the_AT play_NN1 seems_VVZ to_TO
be_VBI an_AT1 attack_NN1 on_II the_AT Church_NN1 ._PUNC

... with **AT**: article; **JJ**: adjective; **NN1**: singular common noun; **IO**: of (as preposition); **VVZ**: -s form of lexical verb; **TO**: infinitive marker 'to'; **VBI**: be, infinitive; **AT1**: singular article; **II**: general preposition; **PUNC**: punctuation

Figure 1: CLAWS horizontal output (word form_POS)

We also applied a Perl program ³ to CLAWS vertical output (cf. Figure 2) to create corpora consisting of lemmas + simplified POS-tags (cf. Figure 3). POS-tags were automatically simplified to match the level of specificity of lemmas, i.e. the six tags available for lexical verbs (VV0, VVD, VVG, VVI, VVN, VVZ) were replaced by a single VV tag.

		POS-tag	Word form	Lemma
0000005	730	AT	The	the
0000005	740	JJ	whole	whole
0000005	750	NN1	point	point
0000005	760	IO	of	of
0000005	770	AT	the	the
0000005	780	NN1	play	play
0000005	790	VVZ	seems	seem
0000005	800	TO	to	to
0000005	810	VBI	be	be
0000005	820	AT1	an	an
0000005	830	NN1	attack	attack
0000005	840	II	on	on
0000005	850	AT	the	the
0000005	860	NN1	Church	church
0000005	870	.	.	PUNC

Figure 2: CLAWS vertical output

the_AT whole_JJ point_NN of_IO the_AT play_NN seem_VV to_TO be_VB an_AT attack_NN on_II the_AT Church_NN ._PUNC
--

Figure 3: CLAWS horizontal output (lemma_simplified POS tag)

We made use of WordSmith Tools 4 (Scott 2004) to create lists of word forms + POS-tags and lemmas + POS-tags for each corpus. In this study, we analyse all lemmas and word forms that were assigned a VV or VV* tag.

3. Verb forms vs. verb lemmas

Any corpus-driven investigation of lexical verbs needs to consider the advantages and disadvantages of using verb lemmas or verb forms as units of analysis. If lemmas are used, the different inflectional forms, eg. *claim*, *claims*, *claimed*, *claiming*, are merged. This is a useful option if the aim of the analysis is to give a general overview of learners' lexical repertoire and/or detect patterns of use that cut across verb forms (e.g. the use of a *that*-clause with the lemma CLAIM). However, as rightly pointed out by Sinclair (1991), lemmas are an abstraction and only using lemmas amounts to losing important information as each word form has its own individual patterning. Sinclair (ibid: 41) sees a future for a new branch of study that focuses on the interrelationships of a lemma and its forms as 'it is not yet understood how meanings

are distributed among forms of a lemma'. He even goes as far as to suggest that lexicographers change the traditional practice of using the 'base' or uninflected form as headword and use 'the most frequently encountered form' instead (ibid: 42), a pioneering view that has so far gone unheeded. In a previous study (Granger & Paquot 2005), we carried out an automatic comparison of a 1 million-word corpus of academic writing and a similar-sized fiction corpus. Using the criteria of keyness, frequency, range and evenness of distribution, we identified 930 lexical items that figured more prominently in the academic corpus than in the fiction corpus. One of the interesting results of the study is that verbs regularly function as EAP keywords in only one or two inflectional forms. As shown in Figure 4, nearly half of the verbs (47%) appear as distinctive EAP items in only one word form and almost a quarter of them (23%) in two word forms. A minority appear in three (19%) or four (or five) word forms (11%).

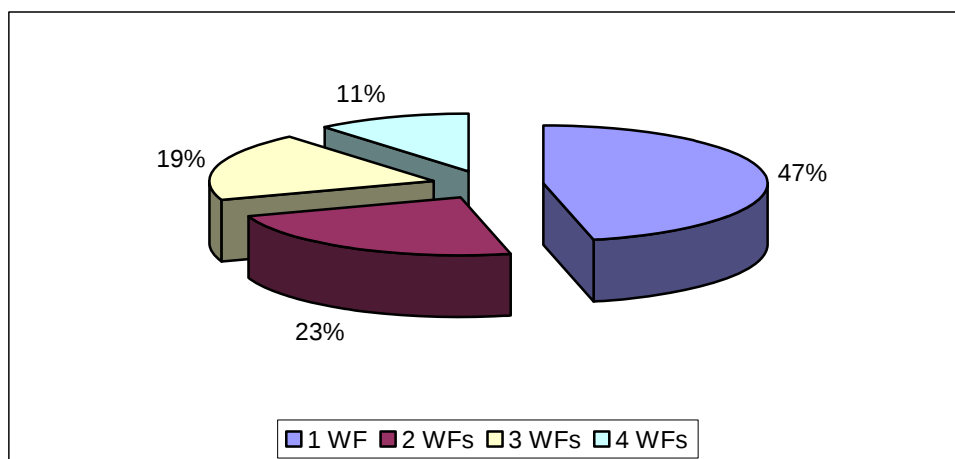


Figure 4: Number of key word forms per key lemma

Table 2 lists some of the verbs in each category. It shows that the verb lemma ASSOCIATE⁴, just like several others such as BASE, CONFINE and LINK, appears as distinctive EAP item in only one word form, i.e. the –ed form. For LACK or COMPRISE,

it is the –ing form that is distinctive and for ENTAIL and REVEAL, the –s form. This shows that, as rightly pointed out by Hyland and Tse (2007: 243), we need to ‘be cautious about claiming generality for families whose meanings and collocational environments may differ across each inflected and derived word form’ (cf. also Oakey 2005). This word of caution has been at the forefront of our analysis of lexical verbs in learner and native writing, the results of which are presented in the following section.

	EAP word forms	EAP lemmas
1 word form	associated based confined linked observed summarized undertaken lacking comprising inducing entails predicts reveals seeks assert benefit coincide participate	ASSOCIATE BASE CONFINE LINK OBSERVE SUMMARIZE UNDERTAKE LACK COMPRISE INDUCE ENTAIL PREDICT REVEAL SEEK ASSERT BENEFIT COINCIDE PARTICIPATE
2 word forms	indicate/indicates amount/amounts conclude/concludes explain/explains emerge/emerges assume/assumes achieve/achieved adopt/adopted specify/specified assess/assessing characterizes/characterized contrasts/contrasting designed/designing	INDICATE AMOUNT CONCLUDE EXPLAIN EMERGE ASSUME ACHIEVE ADOPT SPECIFY ASSESS CHARACTERIZE CONTRAST DESIGN
3 word forms	argue/argues/argued suggest/suggests/suggesting show/shown/shows	ARGUE SUGGEST SHOW

	discuss/discussed/discussing illustrate/illustrates/illustrated	DISCUSS ILLUSTRATE
4 word forms	include/included/including/includes exist/existed/existing/exists develop/develops/developed/developing	INCLUDE EXIST DEVELOP

Table 2: EAP word forms vs EAP lemmas

4. Lexical verbs in learner academic discourse

In this section we draw up lists of the lexical verbs used in ICLE and compare the results with those used in ACAD. We first focus on verb lemmas for the insights they provide into learners' lexical stock of EAP verbs (section 4.1) and then on verb forms (4.2) to uncover new perspectives on learners' preferred and dispreferred EAP patterns (section 4.3).

4.1. EAP verb lemmas

The lists of the top 100 verb lemmas in ICLE and ACAD are included in Appendix 1. Table 3 shows the degree of overlap in the top 100 verbs in each corpus. Of the 148 different verbs, about 35% (N=52) are shared by the two corpora and around a third (32.4%; N=48) are only found in one of the two lists. Among the shared verbs quite a number display marked differences in ranking: WANT (rank 8 in ICLE vs 46 in ACAD), TRY (rank 19 vs 49), HELP (21 vs 66), SHOW (28 vs 9), PROVIDE (40 vs 16).

ICLE only	ICLE and ACAD	ACAD only
AFFECT, AGREE, BAN, BUY, CLAIM, COMMIT, DECIDE, DIE, DREAM, EARN, EAT, ENJOY, FACE, FIGHT, FORGET, GROW, HAPPEN, HEAR, IMAGINE, IMPROVE, KILL, LEARN, LET, LIKE, LOSE, MEET, MENTION, PAY, PLAY, PREPARE, PREVENT, PROTECT, PROVE, READ, REALIZE, SMOKE, SOLVE, SPEND, START,	ACCEPT, ALLOW, ASK, BECOME, BEGIN, BELIEVE, BRING, CALL, CAUSE, CHANGE, CHOOSE, COME, CONSIDER, CREATE, DEVELOP, DISCUSS, EXIST, FEEL, FIND, FOLLOW, GET, GIVE, GO, HELP, INCREASE, KEEP, KNOW, LEAD, LEAVE, LIVE, LOOK, MAKE, MEAN, NEED, PROVIDE, PUT, REDUCE, SAY, SEE, SEEM,	ACHIEVE, ACT, ADD, APPEAR, APPLY, ARGUE, ARISE, ASSUME, BASE, CARRY, COMPARE, CONTAIN, CONTINUE, DEAL, DEFINE, DEPEND, DESCRIBE, DETERMINE, DRAW, ESTABLISH, EXPECT, EXPLAIN, EXPRESS, FORM, HOLD, IDENTIFY, IMPROVE, INDICATE, INVOLVE, MOVE, NOTE, OBTAIN, OCCUR,

STATE, STAY, STOP, STUDY, SUFFER, SUPPORT, TALK, TEACH, WATCH	SHOW, SPEAK, TAKE, TELL, THINK, TRY, TURN, UNDERSTAND, USE, WANT, WORK, WRITE	OFFER, POINT, PRESENT, PRODUCE, RECEIVE, REFER, REGARD, RELATE, REMAIN, REPRESENT, REQUIRE, SET, SUGGEST, TEND, TREAT
---	--	---

Table 3: Top 100 verbs: ICLE vs. ACAD

In addition, many of the top 100 verbs (84.5 %, N = 125) display marked differences in frequency: 55.4 % (N = 82) are overused in ICLE and 29 % (N = 43) are underused. Only 15.5 % (N = 23) are used with similar frequencies. The list of the top 50 underused verb lemmas in ICLE are presented in Table 4 in decreasing order of keyness.

Lemma	Frequency in ICLE	Frequency in ACAD	Log-likelihood
<u>DESCRIBE</u>	273	1080	947.8
<u>OCCUR</u>	324	947	664.5
<u>NOTE</u>	73	527	622.8
<u>SUGGEST</u>	500	1079	558.6
<u>REQUIRE</u>	589	1072	444.5
<u>CONTAIN</u>	233	655	444.3
<u>OBTAIN</u>	310	728	414.4
<u>IDENTIFY</u>	120	471	411.2
<u>INVOLVE</u>	497	939	410.4
<u>ASSUME</u>	186	565	409.4
<u>DERIVE</u>	73	372	377.1
<u>FOLLOW</u>	767	1127	327.1
<u>INCLUDE</u>	468	805	306.8
<u>RECORD</u>	37	252	291.1
<u>DETERMINE</u>	236	531	288.4
<u>REMAIN</u>	555	869	283.5
<u>APPEAR</u>	593	901	278.5
<u>ATTEMPT</u>	69	294	270.1
<u>DEMONSTRATE</u>	98	337	268.7
<u>MEASURE</u>	72	296	266.1
<u>RESPOND</u>	35	224	252.3
<u>ASSESS</u>	34	211	234.6
<u>HOLD</u>	627	881	233.9
<u>PRODUCE</u>	737	979	230.2
<u>ASSOCIATE</u>	144	367	227.2
<u>INTERPRET</u>	71	267	226.6
<u>REPORT</u>	175	403	224.6
<u>GENERATE</u>	81	276	218.6

DEFINE	271	498	209.3
<u>REFER</u>	283	507	205.4
ESTABLISH	391	618	205
RETAIN	52	220	201.2
CONSTITUTE	114	305	197.8
<u>YIELD</u>	17	152	193.4
<u>RELATE</u>	342	554	191.6
COLLIDE	7	126	190
ILLUSTRATE	120	304	187
INDICATE	286	488	183.6
VARY	116	288	173.7
SPECIFY	22	149	171.8
CALCULATE	46	189	169.8
EMERGE	72	226	168.1
<u>ARISE</u>	297	481	166.3
RECOGNIZE	197	373	163.5
<u>EXTEND</u>	141	306	159.4
CONSENT	4	98	155.2
ADD	305	474	152.6
<u>REPRESENT</u>	331	498	151.1
<u>OUTLINE</u>	7	104	151
REMOVE	130	285	150.3
<u>DESCRIBE</u>	273	1080	947.8

Table 4: Top 50 underused verb lemmas in ICLE in decreasing order of keyness

Approximately half (23 / 50, i.e. 45.1 %) of the 50 most underused verb lemmas in ICLE are EAP words according to the AWL. These are printed in bold in Table 4. All the other words except one (COLLIDE) are words from the General Service List (GSL).⁵ If Paquot's AKL is used instead, the proportion of underused EAP words rises sharply to reach a staggering 88 % (44 / 50). As such, the AKL is highly useful in uncovering all the words highlighted by the comparison with the AWL plus a large number of other words, such as DESCRIBE, SUGGEST, NOTE or INCLUDE, which fill important roles in EAP and therefore deserve to be brought to students' attention (AKL words are underlined in Table 4). As most of the verbs are polysemous, a fine-grained semantic classification would require manual scanning of each verb use in context, which clearly falls beyond the scope of this article. However, even without an examination of the verbs in context, the contents of table 4

make it apparent that the majority of the underused verbs fall into three categories: communication verbs (DESCRIBE, SUGGEST, NOTE, DEFINE, RESPOND, REPORT, ADD, SPECIFY); cognition verbs (ASSUME, DERIVE, INTERPRET, ASSESS) and relational verbs (APPEAR, REQUIRE, REMAIN, INCLUDE, INVOLVE).

By contrast, the large majority (45, viz. 90 %) of the top 50 overused words (see Table 5) belong to the General Service List (in bold in Table 5). Besides topic-dependent verbs like DREAM, BAN or SMOKE⁶, the list contains several verbs that are marked by Biber et al (1999) as typical of conversation (e.g. THINK, GET, GO, KNOW, LIKE, WANT) and/or highlighted by Hinkel (2004) as not appearing in EAP texts (e.g. FEEL, LIKE, TRY, WANT). Most are activity verbs (HELP, PUNISH, WORK, TEACH, PLAY) and mental verbs of cognition, perception and affection (THINK, LOVE, FEEL, REALIZE). The list also contains the overused verb of communication SAY. One overused word that is not in the GSL (CREATE) belongs to the AWL but the other four (BAN, IMPORT, RECYCLE, REHABILITATE) are neither in the AWL nor in the AKL. Five overused verbs (STUDY, USE, SOLVE, BECOME, CREATE) appear in the AKL list (underlined in Table 5).

Lemma	Frequency in ICLE	Frequency in ACAD	Log-likelihood
THINK	8711	1331	3245.8
GET	7531	1113	2887.9
DREAM	2453	19	2231
WANT	5169	677	2182.6
WATCH	2331	97	1666
LIVE	4110	578	1641.4
BAN	1358	20	1167.2
LEARN	2768	426	1023.7
PAY	2385	335	953.12
LIKE	2039	266	863.05
GO	5268	1524	837.72
BUY	1464	119	822.81
NEED	3928	1027	753.39
SMOKE	921	24	729.3
SPEND	1732	230	723.28
HELP	2632	555	694.64

TRY	2907	661	693.06
FORGET	1057	83	603.71
KILL	1450	206	574.03
STUDY	1612	268	554.73
PLAY	1963	392	554.25
IMPORT	653	12	546.05
BECOME	5066	1763	521.22
START	1884	391	507.32
EARN	751	40	498.95
KNOW	4941	1742	490.09
FEEL	2530	663	483.2
BELIEVE	2303	582	467.46
TEACH	1243	202	437.15
WORK	2996	917	423.03
SAY	5567	2159	408.81
PUNISH	645	41	402.54
CHANGE	2128	569	392.31
USE	6785	2808	389.66
MAKE	8863	3897	388.57
IMAGINE	970	140	379
FIGHT	944	135	371.66
CREATE	1891	498	358.03
RECYCLE	388	3	352.83
SOLVE	1072	191	343.97
HAPPEN	1621	421	314.32
AFFORD	564	55	288.41
REHABILITATE	311	3	278.27
REALIZE	794	130	277.24
LET	1349	339	276.28
KEEP	1870	571	265.35
LOVE	542	58	262.41
MASTER	331	9	260.05
SAVE	664	96	259.05
EDUCATE	450	36	254.78

Table 5: Top 50 overused verb lemmas in ICLE in decreasing order of keyness

4.2. EAP verb forms

With a view to assessing the relative merits of a lemma vs. word form approach, we replicated the analysis described in the preceding section with verb forms instead of lemmas. While the analysis revealed a wide area of overlap between the two analyses, it also demonstrated that an exclusive focus on lemmas is liable to distort the picture

and hide some major differences between expert and learner use. This distortion can take two different forms: (1) similar frequencies at the lemma level hide over- and/or underuse at the verb form level (cf. Table 6); (2) overuse or underuse at the lemma level affects only some of the verb forms (cf. Tables 7 and 8). A good example of the first type of distortion is the verb CONCLUDE (Table 6), which displays no difference in frequency at the lemma level, but in fact turns out to display an overuse of the infinitive form (*conclude_VVI*) coupled with a significant underuse of the 3rd person singular of the simple present tense (*concludes_VVZ*) and the simple past form (*concluded_VVD*). The second type can be illustrated by the verb ARGUE (Table 7) whose overall underuse at the lemma level conceals an overuse of the simple present form (except for the 3rd person singular) and the verb CAUSE (Table 8) whose overall overuse conceals an underuse of the –ing form.

Lemmas with similar frequency	Overused verb forms	Underused verb forms
ACCESS	<i>access</i> (VVI)	<i>accesses</i> (VVZ)
ALLOW	<i>allowed</i> (VVN)	<i>allowing</i> (VVG)
CONCLUDE	<i>conclude</i> (VVI)	<i>concludes</i> (VVZ), <i>concluded</i> (VVD)
DISCUSS	<i>discuss</i> (VVI)	<i>discusses</i> (VVZ), <i>discussed</i> (VVN)
LEAD	<i>lead</i> (VV0), <i>lead</i> (VVI)	<i>led</i> (VVD), <i>led</i> (VVN)
PROVIDE	<i>provide</i> (VV0)	<i>provided</i> (VVD), <i>provided</i> (VVN)

Table 6: Lemmas vs. verb forms: lemmas with similar frequencies in ICLE and ACAD

Underused lemmas	Underused verb forms	Overused verb forms
SEE	<i>see</i> (VV0), <i>saw</i> (VVD), <i>seen</i> (VVN)	<i>see</i> (VVI)
SHOW	<i>showed</i> (VVD), <i>shown</i> (VVN)	<i>showed</i> (VVN)
ARGUE	<i>argued</i> (VVN), <i>argued</i> (VVD), <i>argues</i> (VVZ), <i>arguing</i> (VVG)	<i>argue</i> (VV0)

Table 7: Lemmas vs. verb forms: underused lemmas

Overused lemmas	Overused verb forms	Underused verb forms
BELIEVE	<i>believe</i> (VV0), <i>believe</i> (VVI)	<i>believed</i> (VVD)

CAUSE	<i>cause (VV0), cause (VVI), causes (VVZ), caused (VVN)</i>	<i>causing (VVG)</i>
FIND	<i>find (VVI)</i>	<i>found (VVN), found (VVD)</i>
GIVE	<i>give (VV0), give (VVI), gives (VVZ), giving (VVG)</i>	<i>given (VVN), gave (VVD)</i>
KNOW	<i>know (VV0), know (VVI), knows (VVZ)</i>	<i>known (VVN)</i>
MAKE	<i>make (VV0), make (VVI), makes (VVZ)</i>	<i>made (VVN)</i>
SEEM	<i>seem (VV0), seems (VVZ)</i>	<i>seemed (VVD)</i>
SPEAK	<i>speak (VV0), speak (VVI), speaking (VVG)</i>	<i>spoke (VVD)</i>
TAKE	<i>take (VV0), take (VVI), taking (VVG)</i>	<i>took (VVD), taken (VVN)</i>
UNDERSTAND	<i>understand (VV0), understand (VVI)</i>	<i>understood (VVN)</i>
USE	<i>use (VV0), use (VVI), using (VVG)</i>	<i>used (VVD), used (VVN)</i>

Table 8: Lemmas vs. verb forms: overused lemmas

It is possible to form a more general picture of the use of EAP verb forms by investigating the breakdown of the different VV tags displayed by the top 100 verb forms in each corpus (full lists in Appendix 2). As shown in Figure 5, the analysis shows striking differences, notably learners' predilection for infinitive forms ($X^2 = 9.9$, $p < 0.01$) coupled with a seeming avoidance of past participle forms ($X^2 = 12.6$, $p < 0.01$).

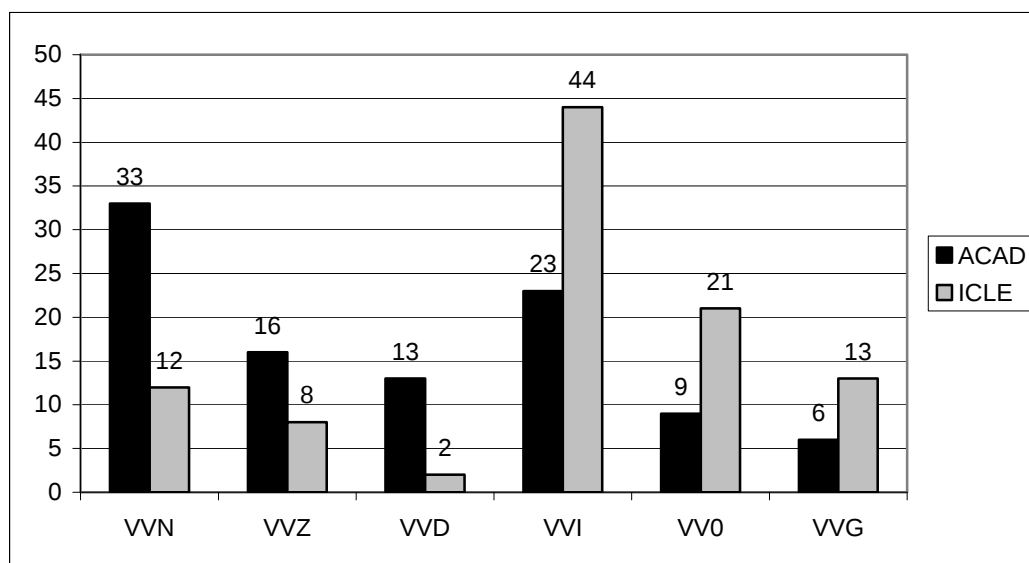


Figure 5: Top 100 verb forms in ACAD and ICLE: breakdown per word category

4.3. Lexico-grammatical patterns of EAP verbs in ICLE

The quantitative differences at the verb form level are indicative of marked differences in phraseological patterning. In order to illustrate this link between verb forms and lexico-grammatical patterns, two representative verbs were selected – CONCLUDE and ARGUE – and submitted to close scrutiny. Concordances proved invaluable in highlighting the patterns of use typical of each corpus.

As shown in Table 6, the lemma CONCLUDE is used with similar frequencies in ACAD and ICLE. However, the word form analysis shows that EFL learners significantly overuse the VVI. This is due to a significant overuse of the connector ‘to conclude’ used in sentence-initial position (130 out of 419 occurrences of the lemma CONCLUDE; 31 %), a use that is very infrequent in ACAD (7 out of 208; 3.4 %). The contrast between the repetitive use of ‘to conclude’ in ICLE and the wider range of patterns used in ACAD appears clearly from the examples (1) to (11).

ICLE

- (1) To conclude we can say that the social position is "gradually" improving. (ICLE-DU)
- (2) To conclude, I will insist on the fact that a professional army is extremely useful (ICLE-FR)
- (3) To conclude I would like to say once more that we have to have faith and hope within ourselves. (ICLE-SW)
- (4) To conclude, I think that the government should ban smoking in restaurants as the health of the public can be improved. (ICLE-CH)
- (5) To conclude, there are neither superior, not inferior cultures. (ICLE-BU)

ACAD

- (6) It must therefore be concluded that the dynamics remains unaltered.
- (7) Finally, the chapter concludes by providing some reflections about the prospects...
- (8) He concludes that the effectiveness of a given system should be based on its ability...
- (9) It is reasonable to conclude from this that, although there are colliding plane wave space-times...
- (10) We may conclude that, in all cases, the opposing waves mutually focus each other...
- (11) We must conclude then that, at the very least, a conditioned inhibitor is a stimulus...

By contrast, the lemma ARGUE is significantly underused by learners: it is almost twice as frequent in ACAD as in ICLE (222 vs 401 per 1 million words). However, as shown in Table 7, this underuse does not affect the base form (VV0), which is overused, due to a recurrent use of the verb ARGUE with *people* and *I* as subject. Here too the contrast between the wide range of patterns displayed by ACAD and the limited range displayed by ICLE is striking (see examples 12 to 34).

ICLE

- (12) Some people argue that television is the greatest invention of the 20th century
(ICLE-PO)
- (13) Some people may argue that the running cost of recycling industry is very high. (ICLE-CH)
- (14) Many people argue that criminals should be punished physically again. (ICLE-GE)
- (15) Quite a few people would probably argue that this is true. (ICLE-NO)
- (16) Another disadvantage is that students perhaps pick up a little bit less when lectured in English, as some people will argue. (ICLE-DU)
- (17) I argue that people should not kill the other people because of discrimination.
(ICLE-TU)
- (18) In this essay, I would like to argue that cohabitation might cause negative consequences, such as knowing a partner's bad personality. (ICLE-JP)
- (19) I would argue that, to a large extent, this is still the case today. (ICLE-SW)
- (20) First I will argue that euthanasia is morally permissible. (ICLE-TU)

ACAD

- (21) It can be argued that experience is in fact the death of innocence.
- (22) It could be argued that this gives the work a sense of coherence.
- (23) It will be argued, however, that the revolution envisaged by Nazi ideology was a failure.
- (24) It has been argued that religion is, in itself, an ideology.
- (25) It is sometimes argued that science textbooks are sexist
- (26) Integration, it is argued, will only work in areas....
- (27) It was argued in Chapter 2 that the criminal law ought to spread its net wider
- (28) Moreover, as argued above, a major reason for having rules....
- (29) In previous chapters I have argued that the decline of this investigatory response....
- (30) Critics have argued that no evidence exists...
- (31) Gergen (1979) also argued that social events are openly competitive.
- (32) In the theatre, he argues, there is an internal dramatist.
- (33) Spinoza shows this by arguing that God is the creator....
- (34) He laid great emphasis on the unity of the Trinity, arguing that root, stem and bark together....

These two examples effectively illustrate the strength of the verb form approach, which functions as a quick way into learners' phraseology. It also shows that over- and underuse need not be taken as negative terms. They can – and indeed should – be taken as prompts for lexical expansion and used with learner populations who wish to attain a native-like mastery of EAP and would benefit from increasing their repertoire of EAP patterns.

5. Issues of text type and domain comparability

Some of the differences between learner and academic writing highlighted in Section 4.3 may be due to differences in text type. As explained in Section 2, ACAD consists of book samples and articles which are expository in nature while the learner texts are short argumentative essays. A large proportion of those verbs that are significantly underused in ICLE (cf. Table 4) perform two essential rhetorical functions in professional academic writing: they are used to quote and report what other scholars have written (e.g. ARGUE, COMMENT, EXPLAIN, NOTE, OBSERVE, PROPOSE, REMARK, REPORT, SUGGEST and WRITE) and to refer to tables, figures and other parts of the text (e.g. DESCRIBE, ILLUSTRATE, SHOW). Learners' underuse of these verbs may thus be (at least) partly explained by a difference in text type as there is no need in argumentative writing to situate one's opinion against what has been written in the literature and typically, argumentative essays do not contain tables and graphs and are too short to include internal reference to chapters and sections. The same explanation may also partly account for learners' underuse of a number of lexico-grammatical patterns, most notably the 'as VVN in/by' structure illustrated in examples 35 to 42.

Quoting and reporting what other scholars have written with the 'as VVN in/by' structure

- (35) If, therefore, the King had turned to Henderson after MacDonald had proffered his resignation, or had sought the views of Labour Privy Counsellors as suggested by Herbert Morrison, he could have been accused of wasting valuable time.
- (ACAD)

- (36) It is also somewhat remarkable that the global structure of the Bell-Szekeres solution, as described by Clarke and Hayward (1989), is very similar to that of the collision of an impulsive gravitational wave with a null shell of matter, as described by Babala (1987). (ACAD)
- (37) The first section of the chapter will provide an overview of patterns and trends in criminal behaviour, as indicated by the official criminal statistics. (ACAD)
- (38) The Christian view of time directed to the future, as presented by St Augustine, differed from the ideas of time current in Classical antiquity in that it was neither cyclic nor would it continue indefinitely without anything essentially new occurring.

Referring to tables, figures and other parts of the text with the ‘as VVN in/by’ structure

- (39) Perhaps because of the symbiotic requirement for such exploitation, phytophagy is restricted within the animal kingdom as shown in Table 6.1. (ACAD)
- (40) In the U.K. less than 3% of the working population are now employed in agriculture, in most advanced countries this figure is now well below 10%, as illustrated in fig. 1.9. (ACAD)
- (41) Thus, the aim with most patients should be to move eventually from assessment and support into a problem-solving approach as described in Chapter 5. (ACAD)
- (42) A possible mechanism whereby the disaggregated and reconstituted bud could form limb-like structures might involve a reaction-diffusion mechanism of the type invented by Turing, as discussed in the previous chapter. (ACAD)

The above provides clear justification for those who argue that learner writing should not be compared with professional academic writing (cf. Hyland and Milton 1997; Lorenz 1998). Indeed, if French learners' essays (ICLE-FR) are compared with American students' argumentative texts (LOCNESS), EFL learners' underuse of these verbs is less marked or nonexistent. Similarly, French learners and American students seem to share a preference for active structures with first person subject pronouns (cf. examples 43 to 52), which may also partly reflect a difference in text type. In argumentative essays such as those contained in ICLE and LOCNESS, 'personal references and subjective attitudes are certainly hard to avoid' (Recski 2004), since essay prompts explicitly encourage learners and native students to give their opinions (e.g. 'Some people say that in our modern world, dominated by science, technology and industrialisation, there is no longer place for dreaming and imagination. What is your opinion?' and 'In the 19th century, Victor Hugo said: "How sad it is to think that nature is calling out but humanity refuses to pay heed." Do you think this is still true nowadays?').

(43) We all know that when intoxicated the response level is lowered. (LOCNESS)

(44) However, today we are beginning to move in a new direction. (LOCNESS)

(45) We cannot change what has happened in history. (LOCNESS)

(46) We must remember, however, that the field of technology is not the only one in which significant advances have been made in the 20th century.

(47) This is a good example of why we need increased gun control (LOCNESS)

(48) I agree that the flag is a sign of heritage for some people. (LOCNESS)

(49) I believe that every woman is beautiful. (LOCNESS)

(50) I feel that the Bible is very real and true. (LOCNESS)

- (51) I know things don't always go our way. (LOCNESS)
- (52) I think most physicians are very sensitive to their patients' wishes regarding death. (LOCNESS)

Both EFL learners and novice writers draw more extensively on high-frequency verbs (e.g. GET, WANT, FEEL, STATE, REALIZE, MAKE, BELIEVE, NEED, THINK) and employ a more restricted number of lexico-grammatical patterns when compared to professional academic writers⁷. EAP-promoted syntactic structures such as the *-ing* supplementary clause and the passive construction are less frequent in LOCNESS and, to a larger extent ICLE, than in ACAD. These differences are indicative of novice writers' – both French learners and American students – imperfect grasp of academic writing conventions (see also Hyland and Milton 1997; Neff et al 2004; Gilquin et al 2007).

Other features of novice writing that appear both in ICLE-FR and LOCNESS are semantic and syntactic errors. For example, both novice writers and EFL learners produce infelicitous cases of 'dangling participles', where the reference is ambiguous (examples 53 and 54).

- (53) *By considering the problem*, a borrower's mindset has begun to take form.
(ICLE-FR) (Correction: *because people are considering the problem*)
- (54) Therefore, *by distributing them in our high schools*, students will be better able to protect themselves and their partners. (LOCNESS) (Correction: *if they are distributed*)

The analysis of LOCNESS, however, reveals that novice L1 writing often appears to occupy an intermediate position between academic writing and EFL learner writing

(cf. Gilquin and Paquot 2008). Novice L1 writers make far fewer semantic and syntactic errors overall than EFL learners. For example, they ‘show a rather balanced use of the reporting verbs *say*, *state*, *show* and *argue*’ (Neff et al 2004), whereas EFL learners underuse the verb ARGUE, which ‘constitutes an important rhetorical device, since it allows the writer to put forward another author’s argument without presupposing its acceptance, either by the writer or the reader’ (Neff et al. 2003: 223).

More importantly, non-native writing contains many examples of problematic language use that are largely learner-specific. French learners sometimes make use of erroneous grammatical collocations and lexico-grammatical patterns with lexical verbs (examples 55 to 59).

- (55) You can discuss **about* several points of view and compare the different opinions. (ICLE-FR) (Correction: discuss several points)
- (56) I think money can claim **being* number one in that matter. (ICLE-FR)
(Correction: claim to be)
- (57) A European cooperation has been attempted but because of administrative constraints has failed. Finally, other fields are difficult to tackle **with* if you are lonely. Among those, I think of the pollution. (ICLE-FR) (Correction: tackle)
- (58) To conclude **with* : it is most likely, I think, that Europe 1992 will see both the birth of a nation and the rebirth of national identity. (ICLE-FR) (Correction: To conclude, it is most ...)
- (59) To conclude **with* this whole debate, I would say that I can hardly find positive arguments to stand for the compulsory military service. (ICLE-FR)
(Correction: To conclude this debate; To conclude or To close the debate)

As put forward by Howarth (1998: 186), '[a] much greater diversity in non-standard phraseology is found in non-native writing, reflecting learners' general lack of awareness of the phenomenon'. EFL learners use lexical verbs in phraseological patterns that are not found in native writing. Example 60 shows that, in ICLE-FR, sentence-initial 'To conclude', is very often followed by a hedging device introduced by a first person pronoun (in italics) while example 61 illustrates the fact that the active frame 'I believe' is repeatedly premodified by an adverb or a modal expression (emphatic 'do', 'would rather') thus creating an impression of overstatement. Similar learner-specific features are described in Lorenz (1998) and Aijmer (2001).

(60) To conclude, *I will say that* we have to be careful...

To conclude, *I would rather consider* Europe as a nation...

To conclude, *my opinion is that* television can be considered as the opium of the masses...

To conclude, *I would say that* science, technology and industrialisation certainly ...

To conclude, *I should say that* in our modern society, new problems and needs...

To conclude, *we should acknowledge that* although television is the new opium...

To conclude, *I would say that* a general authority can be very efficient, but...

To conclude, *I shall once more insist on the fact that* a world in which dreams...

To conclude, *we would say that* Marx was right...

To conclude, *we can say that* many people are today addicted to television.

To conclude, *I will insist on the fact that* a professional army is extremely useful...

- (61) I really believe that every country needs such an army since you can not prevent...

But I really believe that most of the prisoners are not at the right place.

I personally believe that negotiation, long and laborious though it may be, is an alternative...

I deeply believe that our society could improve by paying more attention to all...

.. as far as I am concerned, I truly believe that this task can only be performed by each student individually.

I do believe that prison is not the right place for any criminal.

I would rather believe that this issue has always existed.

In addition, a significant proportion of learner specificities are transfer-related.

French learners map the meaning of French DISCUTER onto the English DISCUSS and use this verb instead of TALK or CHAT (examples 62 and 63). As shown in Granger et al (2006), they produce erroneous verb + noun collocations which have direct translation equivalents in French, e.g. **make abstraction of* (= *faire abstraction de*), **make a step* (= *faire un pas*) and **make part of* (= *faire partie de*) (examples 64 to 66) (cf. also Nesselhauf 2005, Gilquin 2007).

- (62) I am indeed a university student and when I discuss with people around me I often heard this thought that is in fact half true and this constitutes the second

reason which I will develop in this essay. (ICLE-FR) (Correction: when I talk to people)

- (63) As a matter of fact, the picture of people discussing and joking before a fireplace seems to belong to the past. (ICLE-FR) (Correction: chatting)
- (64) No matter if people are unhappy around you! Instead of taking their human feelings into account, what you have to do in business is to try to solve a problem by means of figures, dollars, subsidies, profits... by counting, calculating and making abstraction of data of any other kind. (ICLE-FR) (Correction: disregarding)
- (65) When Man makes a step he wants to go further and he will therefore make a second one. (ICLE-FR) (Correction: takes a step)
- (66) From now on, they may have the feeling that they make part of Europe. (ICLE-FR) (Correction: they belong to)

French learners' repeated use of first person plural forms of the imperative can also be partly attributed to L1 influence (cf. Paquot 2008). The first person plural imperative is a common rhetorical strategy to organize discourse in French formal writing (cf. examples under 67) and French learners seem to transfer its stylistic profile to English (examples under 68).

(67) **French student writing**

Envisageons tout d'abord la question économique.

Examinons quelques exemples pour tenter d'y voir plus clair.

Considérons un instant le cinéma actuel.

Pensons, par exemple, à l'Espagne, qui, pendant quatre à huit siècles, a appris à côtoyer les peuples arabes.

(68) **French learner writing in English (ICLE-FR)**

Let's consider the situation in Belgium.

Let's first have a look at what is Europe actually.

Now let's move on to our third category of criminals.

Let's try to find the most important principles which are urging people to react as they do.

So let us analyse the potential assets of this country...

Let us comment on the second statement: ...

Let us now examine the second solution.

Let us explain these two points.

These results clearly show that learners and novice L1 writers cannot be included in one undifferentiated category of novice writers. The two types of writing have some shared characteristics as both writer populations are students who are learning academic writing conventions. However, there is only a partial overlap between the difficulties of EFL learners and novice L1 writers. Perhaps unsurprisingly, the learner writing contains many examples of difficulties that arise from the fact that the learners are writing in a foreign language (difficulties for example with selecting the appropriate preposition after a verb, or the right verb with delexical nouns such as *claim*, *decision*, and *argument*) and are strongly influenced by their L1.

6. Conclusion

The field of EAP vocabulary has so far been largely dominated by a lemma-based approach. Our study shows that a dual approach – combining both lemmas and word forms – gives us a more precise picture of the diversity of form-meaning mappings

that characterizes the use of EAP verbs and that automatic retrieval of verbs from academic texts produced by EFL learners and expert writers is a powerful first step towards our goal of understanding EFL learner difficulties with a view to enhancing EAP teaching tools. Three main findings emerge from our study. The first is that EFL learners significantly underuse the majority of ‘academic verbs’, i.e. verbs like *include*, *report* or *relate*, that express rhetorical functions at the heart of academic writing, and instead tend to resort to ‘conversational verbs’, i.e. verbs like *think* or *like*, that are characteristic of informal speech. The second is that when learners use academic verbs, they tend to restrict themselves to a very limited range of patterns, which contrasts sharply with the rich patterning that characterizes expert writing. Our study therefore adds support to Ellis et al’s (in press) observation that, even at an advanced proficiency level, learners still “need help to recognize the distinctive formulas that are special to EAP”, a major prompt for including language awareness exercises targeting these formulas in EAP classes. Thirdly, a comparison between ICLE and LOCNESS data has demonstrated that, while novice native writers share a number of problems with EFL learners, the latter are faced with a much wider range of difficulties, many of which are exclusive to the learner population. As a result, blanket EAP textbooks targeting both EFL/ESL learners and novice native writers are bound to be too simplistic. Corpus-based studies like this one demonstrate the tremendous potential of corpus approaches to EAP but also the many challenges they pose to EAP researchers. Besides the issue of comparability which, as shown in our study, has a major impact on the results and the conclusions that can be drawn from them, a series of other issues require further investigation. Prime among these is the very notion of academic prose as a single register, “an overly blunt instrument” according to Hunston (2002: 103), and one whose very existence has been called into question in a number of recent EAP and ESP studies (cf. Hyland 2007). A natural next step in our work is to investigate to what extent the verb patterns displayed in our large academic corpus hold across individual disciplines (cf. Granger & Paquot in preparation). This investigation is just one small step on the long journey to map out the features of native and learner EAP corpora – still a largely underexplored territory.

Footnotes

1. See Ädel (2006: 205-208) for an excellent discussion of corpus comparability in learner corpus research.
2. See <http://ucrel.lancs.ac.uk/claws/>
3. Perl is a programming language that is most helpful to corpus linguists as it provides powerful text processing facilities (cf. Danielsson 2004).
4. From here on, we use small caps to refer to lemmas and italics for word forms.
5. The version of the GSL used is the one that was created in 1995 by John Bauman and Brent Culligan. This list includes all 2,000 capitalized headwords from the original General Service List of West (1953), plus 284 more, ranked and presented in frequency order based on the Brown Corpus. It is available from <http://www.auburn.edu/~nunnath/engl6240/wlistgen.html>.
6. This overuse is clearly topic related: for example, the verb BAN in essays on *‘Citizens in the USA should not be allowed to own guns’*; *‘The role of censorship in Western society’* or the verb DREAM in essays on *‘Some people say that in our modern world, dominated by science and technology and industrialisation, there is no longer a place for dreaming and imagination. What is your opinion?’*
7. The large differences in size between ACAD and ICLE, on the one hand, and ICLE-FR and LOCNESS, on the other, make a fully-fledged comparison in terms of over- and underuse inappropriate. In this section, we therefore focus more on qualitative than purely quantitative differences.

References

- Ädel, A. (2006), The use of metadiscourse in argumentative texts by advanced learners and native speakers of English. Amsterdam: Benjamins.

- Aijmer, K. (2001), '*I think* as a marker of discourse style in argumentative Swedish student writing', in K. Aijmer (ed.), A Wealth of English. Studies in Honour of Göran Kjellmer. Göteborg: Acta Universitatis Gothoburgensis, pp. 247-257.
- Altenberg, B. and Granger, S. (2001), 'The grammatical and lexical patterning of MAKE in native and non-native student writing'. Applied Linguistics, 22, (2), 173-195.
- Biber, D. (1988), Variation across speech and writing. Cambridge: Cambridge University Press.
- Biber D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999), Longman Grammar of Spoken and Written English. Harlow: Longman.
- Britton, B. K. (1994), 'Understanding expository text: Building mental structure to induce insights', in M. A. Gernsbacher (ed.), Handbook of psycholinguistics. New York: Academic, pp. 641-674.
- Burnard, L. (2003), Reference guide for BNC-Baby. Available from <http://www.natcorp.ox.ac.uk/corpus/baby/index.html>
- Charles, M. (2006a), 'Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines'. English for Specific Purposes, 25, 310-331.
- Charles, M. (2006b), 'The construction of stance in reporting clauses: a cross-disciplinary study of theses'. Applied Linguistics, 27, 492-518.
- Coxhead, A. (2000), 'A New Academic Word List'. TESOL Quarterly, 34, (2), 213-238.
- Danielsson, P. (2004), 'Simple Perl programming for corpus work', in J. Sinclair (ed.), How to use corpora in language teaching. Amsterdam: Benjamins, pp. 225-246.

- De Cock, S. (2003), Recurrent Sequences of Words in Native Speaker and Advanced Learner Spoken and Written English: a corpus-driven approach. Unpublished PhD dissertation. Louvain-la-Neuve: Université catholique de Louvain.
- Ellis, N. C., Simpson-Vlach, R., and Maynard, C. (in press), 'Formulaic Language in Native and Second-Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL'. TESOL Quarterly, 41, 3. (Special Issue on Psycholinguistics and TESOL).
- Garside, R. and Smith, N. (1997), 'A Hybrid Grammatical Tagger: CLAWS4', in R. Garside, G. Leech and A. McEnery (eds), Corpus annotation: linguistic information from computer text corpora. New York: Addison Wesley Longman, pp. 102-121.
- Gilquin, G. (2007), 'To err is not all: what corpus and elicitation data can reveal about the use of collocations by learners'. ZAA, 55, (3), 273-291.
- Gilquin, G., Granger S. and Paquot M. (2007a), 'Learner corpora: the missing link in EAP pedagogy'. Journal of English for Academic Purposes, 6, (4), 319-335 (Special issue on Corpus-based EAP Pedagogy)
- Gilquin, G., Granger, S. and Paquot, M. (2007a), 'Writing sections', in M. Rundell (Editor in chief) Macmillan English Dictionary for Advanced Learners (second edition). Oxford: Macmillan Education, pp. IW1-IW29.
- Gilquin, G. and Paquot, M. (2008), 'Too chatty: Learner academic writing and register variation'. English Text Construction, 1, (1), 41-61.
- Gramley, S. and Pätzold, M. (1992) A survey of Modern English. London: Routledge.
- Granger, S. (1996), 'From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora', in K. Aijmer, B. Altenberg and M. Johansson (eds), Languages in Contrast. Text-based cross-linguistic studies. Lund Studies in English 88. Lund: Lund University Press, pp. 37-51.

- Granger, S. (1997), 'On identifying the syntactic and discourse features of participle clauses in academic English: native and non-native writers compared', in J. Aarts, I. de Mönnink and H. Wekker (eds), Studies in English Language and Teaching. Amsterdam and Atlanta: Rodopi, pp. 185-198.
- Granger, S., Dagneaux, E., Meunier, F. and Paquot, M. (forthcoming), The International Corpus of Learner English. Handbook and CD-ROM (second edition). Louvain-la-Neuve: Presses Universitaires de Louvain. Available from <http://www.i6doc.com>
- Granger, S. and Paquot, M. (2005), The phraseology of EFL academic writing: Methodological issues and research findings. Paper presented at AAACL6 & ICAME26, 12-15 May 2005, Ann Arbor, Michigan.
- Granger, S. and Paquot, M. (in preparation), 'In search of General Academic English: A corpus-driven study'. Paper submitted to the International Conference on L.S.P: 'Options and practices of LSP practitioners', 7-8 February 2009, Heraklion, Crete.
- Granger, S., Paquot, M. and Rayson, P. (2006), 'Extraction of multi-word units from EFL and native English corpora. The phraseology of the verb "make"', in A. Häcki Buhofer and H. Burger (eds), Phraseology in Motion I: Methoden und Kritik. Akten der Internationalen Tagung zur Phraseologie (Basel, 2004). Baltmannsweiler: Schneider Verlag Hohengehren, pp. 57-68.
- Hiltunen, T. (2006), 'Coming-to-know verbs in research articles in three academic disciplines', in Proceedings of the 5th International AELFE (Asociación Europea de Lenguas para Fines Específicos) Conference, 246-251. Available from <http://www.unizar.es/aelfe2006/>
- Hinkel, E. (2002). Second Language Writers' Text. Linguistic and Rhetorical Features. Mahwah, New Jersey & London: Lawrence Erlbaum Associates.

- Hinkel, E. (2004), Teaching Academic ESL Writing: Practical Techniques in Vocabulary and Grammar. Mahwah, New Jersey & London: Lawrence Erlbaum Associates.
- Howarth, P. (1998), 'The Phraseology of learners' academic writing', in A.P. Cowie (ed), Phraseology: Theory, Analysis, and Applications. Oxford, Oxford University Press, pp. 161-186.
- Hunston, S. (2002), Corpora in applied linguistics. Cambridge: Cambridge University Press.
- Hyland, K. (1996), 'Nurturing Hedges in the ESP Curriculum'. System, 24, (4), 477-490.
- Hyland, K. (1999), 'Academic attribution: Citation and the construction of disciplinary knowledge'. Applied Linguistics, 20, 341-67.
- Hyland, K. (2008), 'Academic clusters: text patterning in published and postgraduate writing'. International Journal of Applied Linguistics, 18, (1), 41-62.
- Hyland, K. and Milton, J. (1997), 'Qualifications and certainty in L1 and L2 students' writing'. Journal of Second Language Writing, 6,(2), 183-205.
- Hyland, K. and Tse, P. (2007), 'Is there an "Academic Vocabulary"?' . TESOL Quarterly, 41, (2), 235-253.
- Johns, T. and Scott, M. (1993), Microconcord corpus (Collection B). Oxford: Oxford University Press.
- Liao, Y. and Fuyuka, Y.J. (2004), 'Avoidance of phrasal verbs: the case of Chinese learners of English'. Language Learning, 54, (2),193-226.
- Lorenz, G. (1998), 'Overstatement in advanced learners' writing: Stylistic aspects of adjective intensification', in S. Granger (ed.), Learner English on Computer. London and New York: Longman, pp 53-66.

- Meyer, P.G. (1997), Coming to know: studies in the lexical semantics and pragmatics of academic English. Tübingen: Gunter Narr Verlag Tübingen.
- Neff, J., Dafouz, E., Herrera, H. , Martínez, F., Rica, J.P., Diez, M., Prieto, R. and Sancho, C. (2003), 'Contrasting learner corpora: the use of modal and reporting verbs in the expression of writer stance', in S. Granger and S. Petch-Tyson (eds), Extending the scope of corpus-based research. New applications, new challenges. Amsterdam and New York: Rodopi, pp. 211-230.
- Neff, J., Ballesteros, F., Dafouz, E., Martínez, F. and Rica, J.P. (2004), 'The expression of writer stance in native and non-native argumentative texts', in R. Facchinetti and F. Palmer (eds), English Modality in Perspective. Frankfurt am Main: Peter Lang, pp. 141-161.
- Nesselhauf, N. (2005), Collocations in a learner corpus. Amsterdam: Benjamins.
- Oakey, D. J. (2005), 'Academic vocabulary in academic discourse: The phraseological behaviour of EVALUATION in Economics research articles', in E. Tognini-Bonelli and G. Del Lungo Camiciotti (eds), Strategies in Academic Discourse. Amsterdam: Benjamins, pp.169–183.
- Paquot, M (2007), EAP vocabulary in EFL learner writing: from extraction to analysis: A phraseology-oriented approach. Unpublished PhD thesis. Université catholique de Louvain, Centre for English Corpus Linguistics.
- Paquot, M. (2008), 'Exemplification in learner writing: a cross-linguistic perspective', in S. Granger and F. Meunier (eds), Phraseology in Foreign Language Learning and Teaching. Amsterdam: Benjamins, pp. 101-119.
- Recski, L.J. (2004), 'Expressing standpoints in EFL written discourse'. Revista Virtual de Estudos da Linguagem, 2, (3). Available online at

http://www.revel.inf.br/site2007/ed_anterior_list.php?id=3 (last accessed on June 20, 2008).

- Shaw, P. (1992), 'Reasons for the correlation of voice, tense, and sentence function in reporting verbs'. Applied Linguistics, 13, 302-319.
- Sinclair, J. (1991), Corpus, Concordance, Collocation. Oxford: Oxford University Press.
- Sjöholm, K. (1998), 'A reappraisal of the role of cross-linguistic and environmental factors in lexical L2 acquisition', in K. Haastrup and A. Viberg (eds), Perspectives on Lexical Acquisition in a Second Language. Lund: Lund University Press, pp. 209-236.
- Swales, J. (1990), Genre Analysis. English in academic and research settings. Cambridge: Cambridge University Press.
- Swales, J. (2004), 'Then and now: A reconsideration of the first corpus of scientific English'. Ibérica, 8, 5-21.
- Swales J. and Feak, C. (2004), Academic Writing for Graduate Students. Essential Tasks and Skills. Ann Arbor: The University of Michigan Press.
- Thomas, S. and Hawes, T. P. (1994), 'Reporting verbs in medical journals'. English for Specific Purposes, 13, 129-148.
- Thompson, G. and Yiyun, Y. (1991), 'Evaluation in the reporting verbs used in academic papers'. Applied Linguistics, 12, (4), 365-382.
- Tognini-Bonelli, E. (2001), Corpus Linguistics at Work. Amsterdam and Philadelphia: Benjamins.
- Werlich, E. (1976), A text grammar of English. Heidelberg: Ouelle and Meyer.

Williams, I.A. (1996), 'A contextual study of lexical verbs in two types of medical research reports: clinical and experimental'. English for Specific Purposes 15, 3, 175-197.

Appendix 1: Top 100 verb lemmas in ICLE and ACAD

ICLE				ACAD			
	Word	Freq.	Rel. freq. per 1 m word		Word	Freq.	Rel. freq. per 1 m words
1	MAKE	8863	2741.24	1	MAKE	3898	1922.20
2	THINK	8711	2694.22	2	SEE	3003	1480.86
3	GET	7531	2329.26	3	USE	2832	1396.53
4	USE	6785	2098.53	4	TAKE	2829	1395.05
5	TAKE	5810	1796.97	5	GIVE	2682	1322.56
6	SAY	5567	1721.82	6	FIND	2231	1100.16
7	GO	5268	1629.34	7	SAY	2172	1071.07
8	WANT	5169	1598.72	8	SHOW	2084	1027.67
9	BECOME	5066	1566.86	9	BECOME	1763	869.38
10	GIVE	5003	1547.38	10	KNOW	1745	860.50
11	KNOW	4941	1528.20	11	COME	1591	784.56
12	FIND	4131	1277.68	12	SEEM	1564	771.25
13	LIVE	4110	1271.18	13	GO	1525	752.02
14	SEE	4096	1266.85	14	THINK	1333	657.34
15	NEED	3928	1214.89	15	PROVIDE	1328	654.87
16	COME	3363	1040.14	16	CONSIDER	1230	606.54
17	WORK	2996	926.63	17	FOLLOW	1127	555.75
18	SEEM	2941	909.62	18	GET	1113	548.85
19	TRY	2907	899.11	19	WRITE	1083	534.06
20	LEARN	2768	856.11	20	SUGGEST	1081	533.07
21	HELP	2632	814.05	21	DESCRIBE	1080	532.58
22	FEEL	2530	782.50	22	REQUIRE	1073	529.12
23	MEAN	2494	771.37	23	NEED	1027	506.44
24	DREAM	2453	758.69	24	MEAN	1016	501.02
25	PAY	2385	737.66	25	LOOK	989	487.70
26	LOOK	2341	724.05	26	PRODUCE	981	483.76
27	WATCH	2331	720.95	27	LEAD	972	479.32
28	SHOW	2313	715.39	28	OCCUR	947	466.99
29	BELIEVE	2303	712.29	29	INVOLVE	940	463.54
30	CHANGE	2128	658.17	30	WORK	918	452.69
31	CONSIDER	2101	649.82	31	APPEAR	903	445.29
32	LIKE	2039	630.64	32	HOLD	882	434.94
33	PLAY	1963	607.14	33	REMAIN	870	429.02
34	BRING	1956	604.97	34	CALL	856	422.12
35	CREATE	1891	584.87	35	DEVELOP	850	419.16
36	START	1884	582.70	36	CAUSE	835	411.76
37	KEEP	1870	578.37	37	ARGUE	818	403.38
38	CAUSE	1855	573.73	38	INCLUDE	807	397.95
39	LEAD	1785	552.08	39	LEAVE	791	390.06
40	PROVIDE	1774	548.68	40	BEGIN	765	377.24
41	SPEND	1732	535.69	41	ALLOW	739	364.42
42	HAPPEN	1621	501.36	42	OBTAIN	728	359.00
43	STUDY	1612	498.58	43	PUT	716	353.08
44	PUT	1521	470.43	44	FORM	685	337.79
45	SPEAK	1495	462.39	45	WANT	677	333.85

46	DEVELOP	1483	458.68	46	BRING	667	328.91
47	BUY	1464	452.80	47	FEEL	664	327.44
48	KILL	1450	448.47	48	TRY	661	325.96
49	LOSE	1448	447.85	49	CONTAIN	656	323.49
50	UNDERSTAND	1372	424.35	50	DISCUSS	656	323.49
51	TELL	1370	423.73	51	APPLY	651	321.02
52	CHOOSE	1363	421.56	52	SET	641	316.09
53	BAN	1358	420.02	53	BASE	637	314.12
54	LEAVE	1350	417.54	54	EXPLAIN	626	308.70
55	LET	1349	417.23	55	ESTABLISH	619	305.24
56	READ	1325	409.81	56	ACCEPT	597	294.40
57	CALL	1302	402.70	57	BELIEVE	583	287.49
58	TEACH	1243	384.45	58	LIVE	582	287.00
59	INCREASE	1215	375.79	59	ASK	579	285.52
60	TURN	1207	373.31	60	PRESENT	572	282.07
61	AFFECT	1199	370.84	61	CARRY	571	281.57
62	TALK	1178	364.34	62	KEEP	571	281.57
63	ALLOW	1149	355.37	63	CHANGE	570	281.08
64	ASK	1143	353.52	64	ASSUME	565	278.62
65	DECIDE	1123	347.33	65	HELP	555	273.68
66	SOLVE	1072	331.56	66	RELATE	554	273.19
67	FORGET	1057	326.92	67	TURN	548	270.23
68	DISCUSS	1047	323.83	68	UNDERSTAND	544	268.26
69	EXIST	1041	321.97	69	INCREASE	541	266.78
70	AGREE	1031	318.88	70	DETERMINE	531	261.85
71	IMPROVE	1024	316.71	71	TELL	531	261.85
72	WRITE	996	308.05	72	NOTE	528	260.37
73	IMAGINE	970	300.01	73	ACT	521	256.92
74	ACCEPT	945	292.28	74	MOVE	521	256.92
75	FIGHT	944	291.97	75	REDUCE	518	255.44
76	MENTION	939	290.42	76	COMPARE	511	251.99
77	COMMIT	933	288.57	77	CONTINUE	508	250.51
78	SUPPORT	932	288.26	78	REFER	507	250.01
79	SMOKE	921	284.86	79	TEND	506	249.52
80	PREPARE	920	284.55	80	REGARD	505	249.03
81	STOP	919	284.24	81	RECEIVE	501	247.06
82	BEGIN	912	282.07	82	CREATE	498	245.58
83	ENJOY	877	271.25	83	DEFINE	498	245.58
84	REDUCE	844	261.04	84	REPRESENT	498	245.58
85	MEET	838	259.18	85	INDICATE	488	240.65
86	HEAR	830	256.71	86	ACHIEVE	486	239.66
87	EAT	809	250.22	87	ARISE	481	237.19
88	SUFFER	795	245.89	88	DEPEND	480	236.70
89	DIE	794	245.58	89	EXIST	479	236.21
90	REALIZE	794	245.58	90	ADD	478	235.71
91	PROVE	776	240.01	91	IDENTIFY	472	232.76
92	STAY	775	239.70	92	POINT	470	231.77
93	CLAIM	771	238.46	93	SPEAK	470	231.77
94	STATE	770	238.15	94	EXPECT	468	230.78
95	FACE	769	237.84	95	CHOOSE	462	227.82
96	PREVENT	769	237.84	96	DEAL	462	227.82
97	FOLLOW	767	237.23	97	DRAW	458	225.85

98	PROTECT	765	236.61	98	EXPRESS	456	224.87
99	EARN	751	232.28	99	TREAT	450	221.91
100	GROW	745	230.42	100	OFFER	439	216.48

Appendix 2: Top 100 verb forms in ICLE and ACAD

ICLE					ACAD				
	Word	POS	Freq.	Rel. freq. per 1 m words		Word	POS	Freq.	Rel. freq. per 1 m words
1	think	VV0	4867	1505.31	1	made	VVN	1227	605.07
2	get	VVI	3423	1058.70	2	given	VVN	1187	585.34
3	make	VVI	3361	1039.52	3	used	VVN	1162	573.01
4	want	VV0	2797	865.08	4	seen	VVN	1022	503.97
5	take	VVI	2230	689.72	5	make	VVI	995	490.66
6	think	VVI	2228	689.10	6	see	VV0	898	442.83
7	using	VVG	2028	627.24	7	found	VVN	875	431.49
8	go	VVI	2024	626.00	8	taken	VVN	784	386.61
9	say	VVI	2009	621.36	9	seems	VVZ	772	380.69
10	need	VV0	2008	621.05	10	shown	VVN	751	370.34
11	know	VVI	1942	600.64	11	take	VVI	746	367.87
12	find	VVI	1932	597.55	12	known	VVN	720	355.05
13	use	VVI	1878	580.85	13	using	VVG	684	337.3
14	see	VVI	1793	554.56	14	see	VVI	629	310.18
15	seems	VVZ	1756	543.11	15	based	VVN	610	300.81
16	get	VV0	1597	493.94	16	said	VVD	582	287
17	help	VVI	1592	492.39	17	described	VVN	576	284.04
18	live	VVI	1574	486.82	18	find	VVI	530	261.36
19	give	VVI	1518	469.50	19	give	VVI	512	252.48
20	know	VV0	1508	466.41	20	called	VVN	512	252.48
21	dreaming	VVG	1454	449.71	21	came	VVD	506	249.52
22	become	VVI	1336	413.21	22	required	VVN	506	249.52
23	make	VV0	1322	408.88	23	considered	VVN	503	248.04
24	learn	VVI	1305	403.62	24	became	VVD	498	245.58
25	makes	VVZ	1293	399.91	25	say	VVI	480	236.7
26	used	VVN	1271	393.11	26	made	VVD	464	228.81
27	made	VVN	1259	389.40	27	said	VVN	437	215.5
28	believe	VV0	1181	365.27	28	took	VVD	436	215
29	pay	VVI	1181	365.27	29	making	VVG	435	214.51
30	go	VV0	1154	356.92	30	makes	VVZ	434	214.02
31	like	VVI	1143	353.52	31	get	VVI	420	207.11
32	work	VVI	1125	347.95	32	use	VVI	413	203.66
33	mean	VVZ	1095	338.67	33	go	VVI	388	191.33
34	use	VV0	1082	334.65	34	provide	VVI	386	190.35
35	take	VV0	1041	321.97	35	means	VVZ	385	189.85
36	watching	VVG	1039	321.35	36	found	VVD	383	188.87
37	become	VV0	1030	318.57	37	obtained	VVN	380	187.39
38	given	VVN	1012	313.00	38	shows	VVZ	379	186.89
39	say	VV0	1007	311.45	39	know	VVI	370	182.46
40	live	VV0	1002	309.91	40	held	VVN	368	181.47

41	keep	VVI	971	300.32	41	need	VV0	356	175.55
42	feel	VV0	957	295.99	42	went	VVD	350	172.59
43	want	VVI	943	291.66	43	think	VVI	345	170.13
44	feel	VVI	916	283.31	44	associated	VVN	345	170.13
45	buy	VVI	909	281.14	45	make	VV0	343	169.14
46	going	VVG	901	278.67	46	written	VVN	342	168.65
47	see	VV0	891	275.58	47	discussed	VVN	332	163.72
48	said	VVD	876	270.94	48	developed	VVN	330	162.73
49	getting	VVG	865	267.54	49	remains	VVZ	330	162.73
50	taken	VVN	862	266.61	50	show	VVI	327	161.25
51	change	VVI	858	265.37	51	appears	VVZ	326	160.76
52	understand	VVI	850	262.90	52	suggests	VVZ	325	160.27
53	living	VVG	842	260.42	53	regarded	VVN	324	159.77
54	banning	VVG	834	257.95	54	becomes	VVZ	322	158.79
55	going	VVGK	831	257.02	55	know	VV0	316	155.83
56	become	VVN	827	255.78	56	become	VVI	315	155.33
57	try	VV0	818	253.00	57	come	VVI	311	153.36
58	come	VVI	813	251.45	58	taking	VVG	305	150.4
59	working	VVG	812	251.14	59	compared	VVN	304	149.91
60	making	VVG	805	248.98	60	produced	VVN	304	149.91
61	considered	VVN	798	246.81	61	follows	VVZ	301	148.43
62	try	VVI	784	242.48	62	gave	VVD	296	145.97
63	find	VV0	741	229.18	63	takes	VVZ	296	145.97
64	let	VV0	732	226.40	64	say	VV0	291	143.5
65	seen	VVN	729	225.47	65	established	VVN	291	143.5
66	taking	VVG	713	220.52	66	look	VVI	288	142.02
67	gives	VVZ	697	215.57	67	think	VV0	287	141.53
68	bring	VVI	695	214.96	68	help	VVI	285	140.54
69	look	VVI	694	214.65	69	related	VVN	284	140.05
70	like	VV0	689	213.10	70	consider	VVI	283	139.55
71	lead	VVI	687	212.48	71	work	VVI	281	138.57
72	said	VVN	687	212.48	72	gives	VVZ	280	138.08
73	called	VVN	685	211.86	73	left	VVN	278	137.09
74	comes	VVZ	680	210.32	74	saw	VVD	275	135.61
75	solve	VVI	664	205.37	75	seem	VVI	270	133.14
76	give	VV0	663	205.06	76	defined	VVN	268	132.16
77	choose	VVI	663	205.06	77	expected	VVN	268	132.16
78	spend	VVI	658	203.51	78	looking	VVG	266	131.17
79	come	VV0	656	202.89	79	produce	VVI	265	130.68
80	speak	VVI	646	199.80	80	comes	VVZ	263	129.69
81	create	VVI	643	198.87	81	provides	VVZ	263	129.69
82	changed	VVN	642	198.56	82	use	VV0	262	129.2
83	found	VVN	642	198.56	83	began	VVD	261	128.71
84	becomes	VVZ	632	195.47	84	take	VV0	260	128.21
85	need	VVI	630	194.85	85	seemed	VVD	260	128.21
86	study	VVI	625	193.31	86	showed	VVD	260	128.21
87	trying	VVG	622	192.38	87	provided	VVN	258	127.23
88	smoking	VVG	620	191.76	88	presented	VVN	256	126.24
89	improve	VVI	619	191.45	89	argued	VVN	252	124.27
90	based	VVN	619	191.45	90	thought	VVN	252	124.27
91	became	VVD	617	190.83	91	working	VVG	251	123.77
92	dream	VVI	617	190.83	92	followed	VVN	251	123.77

93	wants	VVZ	614	189.90	93	thought	VVD	250	123.28
94	believe	VVI	609	188.36	94	explain	VVI	249	122.79
95	mean	VVI	602	186.19	95	occur	VVI	245	120.82
96	play	VVI	602	186.19	96	depends	VVZ	245	120.82
97	watch	VVI	600	185.57	97	occurs	VVZ	244	120.32
98	needs	VVZ	599	185.26	98	going	VVG	243	119.83
99	seem	VV0	586	181.24	99	involves	VVZ	243	119.83
100	looking	VVG	584	180.63	100	want	VV0	242	119.34