Linguagem & Ensino, Vol. 6, No. 1, 2003 (11-33)

Schoolchildren writing: A corpus-based analysis

(Escrita de alunos: Uma análise baseada em corpus)

Tony BERBER SARDINHA
Pontifícia Universidade Católica de São Paulo
Marilisa SHIMAZUMI
SBCI e Pontifícia Universidade Católica de São Paulo

ABSTRACT: This paper reports a preliminary description of a sample of the APU (Assessment of Performance Unit) archive from a Corpus Linguistics perspective. The APU archive contains thousands of essays and letters written by schoolchildren in Britain. For the purposes of the current investigation, a sample of the handwritten texts was entered into the computer and compared to a corpus of texts written by adults (The Guardian newspaper). The comparison was carried out by computational means using a wide range of techniques, and it brought out some of the typical characteristics of the students' writing.

RESUMO: Este trabalho apresenta uma descrição preliminar de uma amostra do arquivo APU (Assessment of Performance Unit) a partir de uma perspectiva da Lingüística de Corpus. O arquivo APU contém milhares de ensaios e cartas escritas por alunos britânicos. Para os propósitos desta investigação, uma amostra dos manuscritos foi transferida para o computador e comparada com *corpus de textos escritos por adultos (jornal 'The Guardian'). A comparação foi efetuada por meios computacionais através de uma gama variada de técnicas, e expôs algumas das características típicas da escrita dos alunos.*

KEYWORDS: Corpus Linguistics, writing, WordSmith Tools
PALAVRAS-CHAVE: Lingüística de Corpus, escrita, WordSmith Tools

INTRODUCTION

This paper reports a corpus-based description of a sample of the APU ('Assessment of Performance Unit Language Monitoring Project') archive. More specifically, the paper is concerned with describing the sample using quantitative methods commonly employed in Corpus Linguistics (McEnery and Wilson, 1996; Kennedy, 1998), particularly those which would enable us to carry out 'computer-assisted comparative analysis' (Stubbs, 1996: 131). This kind of analysis is needed because 'otherwise we cannot know what is typical or atypical, or whether features of texts are significant, linguistically or ideologically, or not' (Stubbs, 1996: 152).

The text sample which the present study is concerned with is a subset of the APU archive. One of the aims of the project of the APU project was to evaluate the writing skills of children and teenagers in British schools. The children were considered for all purposes native speakers of English, although it is possible that immigrant students for whom English is a second or foreign language may have taken part in the collection. The focus of the research presented here differs from the aim of the original APU project. The goal of the current investigation is to characterize the writing of the young learners as represented in a sample of the APU archive using computational tools.

A computer-based comparative analysis can be implemented through the extraction and interpretation of key words (Scott, 1997, 2000; Stubbs, 1996: 165ff). A key word is a word of unusual frequency, either higher or lower than expected in relation to a reference, which is usually the word frequencies in a corpus that is larger than the text or texts one is interested in. Specific software reads in two word lists and calculates the unusualness based on the frequency counts of each word. The ‘keyness’ is established by means of statistical tests. The resulting key word list is usually interpreted as the words which best characterise the target text in terms of its content. J. R. Firth was one of the first linguists to show an interest in key words. According to him, key words are 'focal' or 'pivotal' terms, whose importance is 'sociological’ (Stubbs, 1996: 165). The computational and the early senses of key words overlap to some degree but are by no means synonymous, since there is no guarantee that the key words pulled out by the computer have any importance beyond the frequency counts. It is during the interpretation phase which follows the extraction of key words that specific kinds of meanings are attributed to the key words.

Another possibility is the comparison of grammatical features across texts. This kind of comparison may offer valuable information about the outstanding features of a text or corpus beyond frequency counts, and involves the tagging of the corpus for part-of-speech. The tagging is typically carried out automatically through specific software ('taggers') which label each word in the texts according to their grammatical class. Frequencies of grammatical features may be compiled and compared across the samples under consideration.

These approaches were used in the course of the investigation reported here. First, the APU corpus was analysed for key words using word frequency counts. Later, it was tagged for part-of-speech and the major grammatical characteristics of the corpus were described.

THE APU ARCHIVE

APU stands for 'the Assessment of Performance Unit (APU) Language Monitoring Project'. This project was funded by the British Government and was conducted by the National Foundation for Educational Research (NFER) for six years, 1979, 80, 81, 82, 83 and 1988. The aim was to survey the attainment in English of British students. The students were in years 6 and 11, that is, 11 and 15 year olds.

Each student who contributed to the archive was asked to write two texts. One was a text about a topic they had a strong opinion on, for example abortion, racial relations, smoking, etc. The other was a job application letter. The students were given a booklet where they would write the compositions. The booklet also included an instruction sheet, which appears in extract 1.

Extract 1: Directions in the booklet

> 'A strong opinion ....
>
> There are many uses for writing apart from describing things, telling stories, writing letters, plays, songs, no-

> notes ... and so on. Writing is also a very useful way of saying what you think about something, and persuading your reader to agree with you. It's a way of making your cas e heard, and getting your viewpoint across. We would like you to think of a subject about which you have a strong opinion. The choice of subject is up to you. Write about it in some detail so as to persuade somebody who does not share your opinion to change their mind, and see things your way. [APU Booklet 2, 1988]

COLLECTING A SAMPLE FOR ANALYSIS

The whole APU archive consists of 60,000 compositions. There were no resources available for turning the whole archive into computer-readable format, and so a small sample had to be drawn from the archive and entered into the computer for the purposes of the current investigation. The decision was taken to draw a sample to reflect the texts written by the older students, namely 15 year olds. This is because it was felt that this age group would perhaps be in an early transition period into adult life and therefore this would warrant an interesting contrast with an adult variety as represented by the reference corpus. In addition, the writing of the 15 year olds would presumably exhibit a higher level of complexity thus allowing for a fairer comparison with the reference corpus.

There were two problems during the collection of the sample from the archive. The first one derived from the fact that the compositions were handwritten, and therefore they could not be scanned. The texts had to be typed one by one instead. This put a limit on the number of texts which could be entered into the computer. Typing hundreds of texts is a tedious time-consuming activity in itself, but in this particular case the task of typing was not made any easier either by the fact that the compositions had all sorts of errors which needed to be maintained in the typed versions. Hence, the typing had to be done very carefully in order to keep those errors, which ended up slowing down the whole process even more. The second problem related to the regulations surrounding the storage of the archive which prohibited any part of the archive from leaving the building where it is kept. This meant that each composition had to be photocopied and typed in a separate place, or alternatively typed in the archive room designated for consultations. The latter option would have been preferred were it not for the fact that there were no computers in the consultation room, nor was a portable computer made available to the researchers. Thus, the only option was to have the compositions photocopied, take them away, and then type each one. Both of these factors ultimately had the effect of reducing the size of the machine-readable sample extracted from the archive.

The final sample (henceforth the APU corpus) comprised 100 texts written by 15 year olds. Of these, 70 were essays and 30 were letters. The total size of the corpus in words is about 29 thousand words (see table 1).

Table 1: The APU corpus

Text type	Texts	Total words	Words per text
Essays	70	25,602	366
Letters	30	3,762	125
Total	100	29,364	294

The corpus is very small in comparison to most present-day corpora. However, two points must be addressed with respect to the issue of corpus size and representativeness. Firstly, as Leech (1991) argues, size is not all-important, mainly when texts are not easily available. In such cases, practical constraints override theoretical considerations. This was clearly the case in the research reported here. Biber et al (1996) used a corpus of a few thousand words to investigate errors in ESL compositions. Secondly, the usefulness of small corpora is to serve as a sample of a specific kind of text for a particular kind of investigation. A small corpus such as the one used here cannot be taken as a representative sample of the writing of 15-year-old schoolchildren in Great Britain. Its purpose is rather to help indicate possible characteristics of the writing of a part of the universe of writers represented in the APU archive. It must be remembered that the APU itself is not a perfect representative of the universe of 15-year-old schoolchildren in Great Britain and therefore even if it were wholly machine-readable, one would still have to be cautious about generalizating from findings obtained from it.

ANALYSING THE CORPUS

The APU corpus was analysed in four ways, namely through the inspection of word frequency comparisons, extraction of lexical phrases (n-grams), word frequency comparison with a reference corpus (key words), and identification of verb patterns. Other procedures could have been selected, such as lemma-token regression analysis, Multidimensional analysis, and lexical density, to name a few, but these demanded a level of statistical expertise that was beyond the ability of the researchers at the time.

Word frequencies

The first task in analysing the corpus was the extraction of word frequencies. The top frequencies are displayed in table 2. As usual, the most frequent items are non-content (function) words. This reflects a structural characteristic of the language whereby non-content words are used more frequently than content ones. Among the lexical words in the frequency list, some words stand out as particularly interesting. For example, 'people' and 'animals', which are the 16th and 26th most frequent words, respectively. It appears that these items are indicative of the recurring themes in texts, such as 'abortion', 'animal cruelty', and 'smoking' which are some of the issues the students had a ‘strong opinion on’.

Lexical phrases

In addition to frequencies of isolated words, a listing of frequencies of lexical phrases was also compiled. Lexical phrases are commonly regarded as revealing of the way the messages in the text are organized. They serve a range of purposes, from discourse markers to interpersonal adjuncts (Nattinger and De Carrico, 1992). They also function as an index for features such as informality, impersonality, and affect. Two kinds of lexical phrases were produced, one for bigrams (pairs) and another for trigrams (three-word combinations), which appear in table 3 below.

Table 2: Word frequencies in the APU corpus

Rank	Word	Freq	%	Rank	Word	Freq	%
1	The	1217	4.1	16	People	289	1.0
2	To	877	3.0	17	As	242	0.8
3	And	873	3.0	18	You	229	0.8
4	A	694	2.4	19	Not	226	0.8
5	I	688	2.3	20	Would	221	0.8
6	Of	549	1.9	21	This	214	0.7
7	Is	464	1.6	22	If	209	0.7
8	They	460	1.6	23	On	203	0.7
9	In	445	1.5	24	But	181	0.6
10	It	404	1.4	25	Or	168	0.6
11	Are	375	1.3	26	Animals	161	0.5
12	Be	351	1.2	27	With	155	0.5
13	For	343	1.2	28	Do	154	0.5
14	Have	329	1.1	29	Them	154	0.5
15	That	305	1.0	30	There	143	0.5

Table 3: Lexical phrases

Pairs	Freq.	Triplets	Freq.
I Am	106	I Think That	24
In The	104	A Lot Of	22
They Are	101	Cruelty To Animals	15
It Is	94	Be Able To	13
Of The	85	I Do Not	13
I Think	79	I Would Be	13
Should Be	75	It Would Be	11
I Have	73	In My Opinion	10
To Be	68	In The Home	8
Would Be	64	Have An Abortion	7
For The	62	If You Are	7
I Would	57	Stay At Home	7
This Is	50	The Fact That	7
And I	48	A Group Leader	6
On The	48	Are Just As	6
Is A	45	I Think It	6
Have A	39	Is In The	6
Think That	38	It Is A	6
Do Not	34	Place Is In	6
If They	34	They Want To	6

There is a considerable number of phrases beginning with ‘I’ in both lists. For example, 'I think' appears 79 times and 'I think that' 24 times. This phrase seems indicative of the corpus, in that the majority of the texts are about what the pupils 'think' about a certain controversial topic. Other phrases formed with 'I', such as 'I am', 'I think', 'I have', 'I would', 'I do not', 'I would be' , and 'I think it' indicate a high level of personal involvement in the issues on the part of the writer, contributing to impart a personal tone to the texts.

Key word analysis

A key word is a word whose frequency is unusual in comparison to a reference corpus (Scott, 1997). The key words, as used here, were computed by the KeyWords program in WordSmith Tools (Scott, 1996). The program compares the frequencies of individual words in the target corpus (in our case, the APU corpus) to those in a reference corpus (the Guardian corpus). A word which is more frequent in the target text than in the reference corpus will typically be key (Scott, 1996). Hence what is meant by key word here is not ‘important word’, since importance is a subjective criterion which depends on qualitative interpretation on the part of the reader or the analyst. A word which is as frequent in the target text as in the reference corpus might be found important by readers of the target text, but would not be a key word for the computational procedure in that specific situation. A word of frequency 1 would not normally reach significance in the statistical tests carried out by the key word procedure and therefore would not be chosen as key, but human readers might find it a key word in the text. The KeyWords program is then simply a tool to help the analyst, but it will not do the analysis for them.

The statistical procedure used by the program to identify the key words was chi-square; more recent versions of the program used log-likelihood (Dunning, 1992). It must be said that the there are other statistical procedures than key words for selecting words based on their frequency, such as Kita's 'cost criteria', but these were not an option in this study because WordSmith Tools, the software package employed in the analysis, did not offer these tests, and also because the key words procedure had already been used in the literature yielding interesting results (Scott, 1997).

The key word procedure provides a powerful technique for contrasting texts and highlighting possible features of interest in the target text or corpus, in spite of the potential differences between the kinds of key words pulled out by the computer and those likely to be suggested or perceived by readers. Some of the kinds of questions that procedure can help answer are 'how is text 1 different from or similar to text 2?' and 'what are the possible topics being discussed in text/corpus A as opposed to text/corpus B?' Again, it is up to the analyst to interpret the key words in view of the goals of the comparison.

A corpus of texts written by adults was used as a reference in the analysis for key words, namely a collection of Guardian newspaper articles published between 1991 and 1994 amounting to 95,075,857 words. Reference corpora such as the one used here are an obligatory element in a key words analysis through the KeyWords program. By comparing the APU texts with the Guardian corpus, it was expected that the key words would reveal the lexical differences between the writing of schoolchildren and adults. A reference corpus, as used here, is different from the better known 'comparison' corpus, since the former is not the focus of the study.

Key words identified through WordSmith's KeyWords program can be of two types: positive and negative. The former are those which occur in the target corpus more often than expected statistically, and the latter are those which are present in the target texts less often than expected. The choice was made to consider positive key words only, since these indicate which words were used by the students more than by the adult writers.

Word lists were created for each corpus and later processed with the KeyWords program. The key words extracted appear in table 4 below. All of the key words have significant chi-square values at p<.05. The table shows the words in order of 'keyness', that is, words higher up are those whose relative frequencies are higher in the APU than in the Guardian To illustrate, the word that is most key is I, which appears 2.3% of the time in the APU and just 0.3% in the Guardian, that is, it is nearly 8 times more common in the APU.

Table 4: Key words extracted by comparing APU to Guardian (GUA)

Word	% in APU	% in GUA
I	2.3	0.3
People	1.0	0.1
Am	0.4	0.0
Think	0.5	0.0
They	1.6	0.4
You	0.8	0.2
Because	0.5	0.1
Do	0.5	0.1
Should	0.5	0.1
Are	1.3	0.4
Children	0.3	0.0
If	0.7	0.2
Women	0.3	0.0
Very	0.4	0.1
Job	0.2	0.0
Them	0.5	0.1
Feel	0.2	0.0
Would	0.8	0.2
Get	0.3	0.1
Have	1.1	0.5
My	0.4	0.1
Child	0.1	0.0
Be	1.2	0.6
Like	0.4	0.1
Some	0.4	0.1
Can	0.4	0.1
Your	0.2	0.0
It	1.4	0.7
Many	0.3	0.1
Say	0.2	0.1
Men	0.2	0.0
Or	0.6	0.2
Young	0.2	0.0
So	0.4	0.2
Just	0.3	0.1
Lot	0.1	0.0
Why	0.2	0.0
Life	0.2	0.1
Don't	0.2	0.0
Go	0.2	0.1
This	0.7	0.4
Thing	0.1	0.0
Not	0.8	0.4
Live	0.1	0.0
Me	0.2	0.1
Work	0.2	0.1
These	0.2	0.1
Is	1.6	1.0
Also	0.3	0.1
School	0.1	0.0
There	0.5	0.2
Give	0.1	0.0
And	3.0	2.2
Try	0.1	0.0
Bad	0.1	0.0
Then	0.2	0.1
About	0.4	0.2
Doing	0.1	0.0
Able	0.1	0.0
All	0.4	0.2
Things	0.1	0.0
Need	0.1	0.0

Table 4: Key words extracted by comparing APU to Guardian (GUA) (Cont.)

Word	% in APU	% in GUA
Know	0.1	0.0
Look	0.1	0.0
Can't	0.1	0.0
Age	0.1	0.0
Black	0.1	0.0
Make	0.2	0.1
Done	0.1	0.0
Take	0.2	0.1
Really	0.1	0.0
See	0.1	0.1
When	0.4	0.2
Help	0.1	0.0
Put	0.1	0.0
Woman	0.1	0.0
Working	0.1	0.0
Good	0.2	0.1
Keep	0.1	0.0
Going	0.1	0.0
Strong	0.1	0.0
Parents	0.1	0.0
Something	0.1	0.0
Want	0.1	0.0
Out	0.3	0.2
Looking	0.1	0.0
Could	0.2	0.1
Test	0.1	0.0
Leader	0.1	0.0
As	0.8	0.6

Some of the lexical words which seemed to be striking in the frequency list turned out to be key words. The key words thus lend statistical support to the perception of those words as indicative of the APU corpus. Nevertheless, the key word lists give prominence to words which were so striking in the frequency list. These words seem to suggest typical themes in the teenager texts, such as ‘family’, ‘women’, ‘school’, and ‘children’. Words which one might expect to come out as top key words such as ‘cruelty’ or ‘abortion’ did not do so because they were apparently just as common in newspaper stories as in the corpus. However, the angle followed by the young writers on those topics seems to differ from that encountered in other situations. The teenage students seem to favor the inclusion of participants that are close by rather than institutions and people which are more distant such as the ‘government’, ‘ministers’, or ‘the Parliament’.

Verb Patterns

A verb pattern consists of a key verb (a key word which was a verb) plus its collocates (words or groups of words) or colligates (word classes). The key verbs in the corpus were: ‘have’, ‘think’, ‘get’, ‘give’, and ‘put’. This analysis was based on the work of Francis and Hunston (1996), which provides a comprehensive inventory of the verb patterns of English based on a large corpus of British English. The patterns listed in their book may be taken to be representations of the actual ways in which verbs are used in real language as attested in a representative corpus. Importantly, they described the patterns in terms of their frequency in the language, which may serve as a parameter for comparison between a sample of a given variety and the English language as a whole. Accordingly, the patterns in the APU corpus may be compared to the patterns for the English language as attested in Francis and Hunston

(1996) to see to what extent the usage of verbs by schoolchildren differs from the typical usage in English.

In order for the verbs patterns to be identified, the APU corpus was tagged for part of speech, with the Birmingham Part of Speech tagger. The frequencies of the individual word classes appear in the appendix. A large proportion of the words in the corpus were verbs (joint total of 20.3%). The decision was then taken to describe the patterns of key verbs (verbs which were also key words). As mentioned above, the usefulness of the analysis for grammatical features lies in indicating how the key words were used in text. The patterns were extracted by running concordances for each key verb and noting down its collocates and colligates.

Table 5: Verb patterns for key words in the APU corpus

	In APU		In Francis and Hunston	
Verb	Pattern	% of total for verb	Pattern	Rank
Have	+Deteminer	24	V n	1
	+ to	29	Phr-modal	1
Think	+ That	27	V That	1
	+ Pronoun	22	V that-deleted	1
Get	+Determiner	23	V n	1
Give	+ Pronoun	58	V n	1
Put	+ Preposition	47	V n Prep	1

Table 5 shows the patterns for the key verbs. Some verbs had more than one pattern associated with them. The 'patterns' column shows the word classes that typically follow each verb. For instance, the most common pattern for ‘have’ is ‘have + determiner’ which accounts for 24% of the usages of ‘have’ in the APU. The table also presents the corresponding pattern in Francis and Hunston (1996). ‘Have + determiner’, for example, appears in Francis and Hunston (1996) as a ‘V n’ pattern (verb + noun). Significantly, Francis and Hunston (1996) also offer the rank of each pattern within their corpus and the COBUILD dictionary. The ‘V n’ pattern, according to them is the most frequent pattern in the English language, hence the number 1 in the table under ‘frequency’. There is a correlation in the rankings for patterns as shown in Francis and Hunston (1996) between frequency and complexity. Accordingly, the most frequent patterns are the least complex ones. What is most striking about the information in table 5 is that all of the patterns for the key verbs in the APU are the most simple in English. This amounts to a major characteristic of the corpus, namely the use of simple verb patterns.

CONCLUDING REMARKS

The four different kinds of corpus-based analyses shown above provided a means for looking at the corpus from different angles. The word frequencies indicated *which* topics seemed to have been written about (abortion, animal care, people’s actions and opinions). The word combination analysis showed how these words were being used in larger units, hence providing more context to illuminate *how* the topics were being written about. The word combinations indicated that writers kept a personal focus on those issues (I am, they are, I have, I do not, they want to, etc.). The key word analysis picked up other important aspects of the writing of the schoolchildren through a comparison with the adult writers. The key words stressed that a personal stance ('I' as the main key word) was a major characteristic of the children's writing, but it also revealed a preference for explaining one's opinions ('because'), and it suggested that the main topics were combined with more local references ('children', 'school', 'young', 'parents', etc.). Finally, the verb pattern analysis showed that the verbs used by the schoolchildren were mostly the most basic patterns in English (a verb followed by a noun, 'that', or preposition).

The text below is a composition from the APU corpus and illustrates how these characteristics highlighted but the different analyses were present in a single text:

> 'I don't think there is any need for violence on television because if a film or programme didn't contain violence it would be just as enjoyable. I resently [sic] saw a film called Robocop and contained scenes which were so violent that it may offend some people but if it had none or very little violence it would be just as good. Also there is no need for all the bad language which are in films. Now a 15 year old can go to a video shop and hire a film which contains bad language and violence which I only think is suitable for an adult. [...] If people who run tv [sic] must screen films which contain violence they should put them on at suitable time say after midnight. I feel very strongly about this and I think something should be done about it.'

The text shows the writer taking a personal stance ('I don't think', 'I recently saw', etc) towards the topic (violence, TV). The text treats the topic by associating it with how this may affect others ('people'). Several key words are present ('I', 'people', 'should', etc), as well as word combinations ('I think', 'should be', 'would be'). The text also uses 'V-n' verb patterns ('contain violence', 'say a film', 'contains bad language', etc., although not the ones formed with key word verbs) as well as 'V + that-deletion' ('think something should be done').

The research reported here cannot provide a definitive profile of the APU, since the corpus analysed is just a small sample from the whole archive. It can, however, indicate that the application of corpus and text analysis methodologies can bring out several aspects of the texts which might otherwise remain unnoticed, or which would be impossible to notice in a manual analysis. Future research could use a larger sample from the APU archive to validate or challenge our findings.

REFERENCES

BIBER, D. ET AL. Corpus linguistics and language teaching: Concordancing and beyond. Colloquium presented at 30th TESOL Convention, March 28, 1996, Chicago, Ill, USA, 1996.

DUNNING, T. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, v. 19, p. 61-74, 1992.

FRANCIS, G. & S. HUNSTON. *Grammar Patterns 1: Verbs.* London: HarperCollins, COBUILD, 1996.

LEECH, G. The state of the art in corpus linguistics. IN: K. AIJMER & B. ALTENBERG (orgs.). *English corpus linguistics - Studies in honour of Jan Svartvik.* London: Longman, 1991.

KENNEDY, G. *An introduction to Corpus Linguistics*. New York: Longman, 1998.

MCENERY, T. & A. WILSON *Corpus Linguistics*. Edinburgh: Edinburgh University Press, 1996.

NATTINGER, J. R. & J. S. DECARRICO. *Lexical Phrases and Language Teaching*. Oxford: OUP, 1992.

SCOTT, M. *WordSmith Tools.* Software for text analysis. Oxford University Press, Oxford, 1996.

----- PC Analysis of key words - and key key words. *System,* v. 25, p. 233-245, 1997.

----- Mapping key words to problem and solution. IN: M. SCOTT & G. THOMPSON (orgs.). *Patterns of Text? In Honour of Michael Hoey*. Amsterdam: John Benjamins, 2000.

STUBBS, M. *Text and Corpus Analysis -- Computer-Assisted Studies of Language and Culture*. Oxford: Blackwell, 1996.

APPENDIX

Word classes in the APU corpus, tagged by the Birmin gham tagger

Tag	Class	Freq.
NN	Noun sing/mass	12.5%
IN	Prep or sub conjunction	12.2%
DT	Determiner	9.3%
PP	Personal pronoun	8.0%
NNS	Noun plural	7.0%
VB	Verb base form	6.9%
RB	Adverb	5.8%
JJ	Adjective	5.6%
CC	Coordinating conjunction	4.1%
VBP	Verb non-3rd sing pres	4.0%
TO	Infinitive marker	3.0%
MD	Modal	2.8%
VBZ	Verb 3rd sing pres	2.6%
VBN	Verb past participle	2.4%
VBG	Verb gerund/pres part	2.3%
NP	Noun proper singular	2.1%
VBD	Verb past tense	2.1%
PPO	Pronoun possessive	1.3%
CD	Cardinal number	1.0%
WRB	Wh-adverb	0.7%
JJR	Adjective, comparative	0.6%
WP	Wh-pronoun	0.6%
WDT	Wh-determiner	0.5%
EX	Existential `there'	0.4%

RP	Particle	0.4%
UH	Interjection	0.4%
FW	Foreign word	0.3%
JJS	Adjective, superlative	0.3%
SYM	Symbol	0.2%
RBR	Adverb, comparat	0.2%
NPS	Proper noun, plural	0.2%
RBS	Adverb, superlative	0.1%
PDT	Predeterminer	0.1%

Recebido: Outubro de 2001
Aceito: Março de 2002

Endereço dos autores:
LAEL/PUC-SP
Rua Monte Alegre, 984
Perdizes
05014-001 - São Paulo, SP

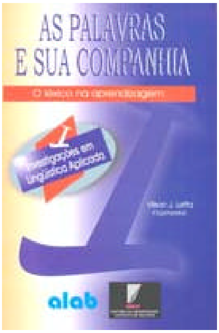
Publicações da EDUCAT

(Em convênio com a ALAB)

Fones: (0xx) 53 284.8236 / 284.8297 - Fax (0xx) 53 225.3105

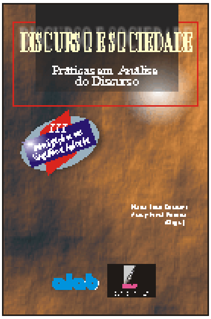
e_mail: educat@phoenix.ucpel.tche.br

site: www.ucpel.tche.br/oficial/educat/comprar.htm



Palavras e sua companhia
O léxico na aprendizagem
Vilson J. Leffa (org.).

Aquisição de Língua Materna e de Língua Estrangeira: Aspectos fonético-fonológicos
Carmen M. Hernandorena (Org.),



Discurso e sociedade: Práticas em Análise do Discurso
Maria José Coracini, Aracy Ernst Pereira (Orgs)

Texto situado: Textualidade e função comunicativa
Leci B. barbisan, Maria Eduarda Giering e Marlene Teixeira (orgs)