BELT+
Brazilian English Language Teaching Journal

ISSN 2178-3640

Porto Alegre, December 2015, v. 6, sp. n. - suppl., s21-s33

http://dx.doi.org/10.15448/2178-3640.2015.s.21311

ORIGINAL ARTICLE

Compilation of a University Learner Corpus

Deise Prina Dutra[1], Andressa Rodrigues Gomide[1]

[1] Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brasil.

ABSTRACT

Corpus Linguistics (CL) and Second Language Acquisition (SLA) areas have been complementary background for researchers interested in contrastive interlanguage analysis (Granger, 1998), shedding light on our understanding of English acquisition by various learner groups. In Brazil there is a paucity of research that describes university learner English so that pedagogical interventions fit their needs. The main objective of this paper is to describe the compilation of a Brazilian university level learner corpus, *CorIsF-Inglês*, and illustrate how a frequency analysis can reveal learner choices when they perform different written tasks. The type of task, independent or integrated, is likely to have influenced the frequency of nouns, verbs and adjectives learners used.

KEYWORDS: Learner corpus; Corpus design; Independent task; Integrated task.

A Criação de um *Corpus* de Aprendizes Universitários

RESUMO

As áreas de Linguística de *Corpus* (LC) e de Aquisição de Segunda Língua (ASL) têm sido pano de fundo complementares para pesquisadores interessados em análise contrastiva da interlíngua (Granger, 1998), iluminando nossa compreensão sobre a aquisição de inglês por aprendizes de vários grupos. No Brasil, há poucas pesquisas que descrevem o inglês de aprendizes universitários que permitam que intervenções pedagógicas sejam adequadas a suas necessidades. O objetivo principal deste artigo é descrever a compilação de um corpus de aprendizes brasileiros, *CorIsF-Inglês*, e ilustrar como uma análise de frequência pode revelar as escolhas dos aprendizes quando eles fazem tarefas de escrita. O tipo de tarefa, independente ou integrada, pode ter influenciado a frequência de substantivos, verbos e adjetivos que os aprendizes utilizaram.

PALAVRAS-CHAVE: *Corpus* de aprendiz; Desenho do *corpus*; Tarefa independente; Tarefa integrada.

Corresponding Author:
DEISE PRINA DUTRA
<deisepdutra@gmail.co>

1. INTRODUCTION

Studies on learner corpus started at the end of the 1980´s (Granger, 2015). Before that, mainly in the 60's and 70's, with error analysis studies, learner language research focused on data that was rarely controlled by testing or language classroom condition (Granger, 1998). Researchers interested in language acquisition were, at times, very concerned with attesting a theory. White (1989) and Schachter (1988), for instance, were concerned in proving the availability of Universal Grammar (Chomsky, 1981) to second language acquisition (SLA). Scholars that sided with cognitive accounts of SLA have argued that language learning involves all aspects of cognitive processing (MacWhinney, 1987) and that interlanguage (Selinker, 1972), learners' production in a second language (SL), should be studied as a system in itself. This cognitive view of learners' use of a SL has matched the interest of some corpus linguists who have been studying learners´ language since the 90´s. The advent of more accessible computers and ways of storing data have made it possible, then, for corpus linguistic tools to be used by more researchers and, according to Granger (2015), contrastive interlanguage analysis (CIA) has flourished since 1996.

Although the increase of CIA studies has been steady, Leech (1998, p. xvi) predicted that when "SLA meets corpus linguistics" such encounter would not be so smooth. Corpus linguists (CL) may not be well prepared in SLA issues and SLA linguists may not be aware of all the tools that corpus linguistics can provide, neither want to focus their investigations on learner's production, but on their mental process in learning a language. Fortunately, Granger's 1998 book was the first volume of its kind, opening the doors for many other publications involving CL and SLA. The growing interest in empirical studies have given support to corpus linguists that study learner production either oral or written.

The compilation of a learner corpus is a challenging issue, especially due to the complexity of collecting a large amount of data. One way of facing this problem is to develop international projects. Some of these projects have been very successful, such as the *International Corpus of Learner English*, (ICLE), which is already in its second version (Granger et al., 2009)[1], and LINDSEI (*Louvain International Database of Spoken English Interlanguage*)[2]. ICLE has been compiled in 16 different countries (e.g. Japan, Belgium, Netherlands, etc.) among university language level students who wrote argumentative essays. LINDSEI also has the same type of participants, yet, data is being collected to form an oral corpus. This project has 20 partners from different countries, (e.g. Greece, Italy, China, etc.) and 13 of them have already completed the data compiling. Although there is a need for more studies on Brazilian learner corpora, some systematic investigations have been done on lexical bundles in written production (Dutra; Berber-Sardinha, 2013; Shepherd, 2009) and on the design of an oral learner corpus (Mello et al., 2012). Other specialized learner corpora have been compiled, such as the Corpus

[1] ICLE's coordinator is Dra. Sylviane Granger (Université catolique de Louvain). Br-ICLE, the Brazilian *subcorpus* of ICLE, is coordinated by Dr. Tony Berber Sardinha from PUC-SP (Berber-Sardinha, 2001). Recently, this *subcorpus* reached 200,000 words and should be part of ICLE in the near future.

[2] LINDSEI's coordinator is Dr. Fanny Meunier (Université catolique de Louvain). LINDSEI-BR is coordinated by Dr. Heliana Ribeiro Mello from UFMG and it is still being compiled.

of Academic Learner English (CALE)[3], which comprises seven different academic text types (e.g. research papers, reading reports, abstracts, reviews, etc.) and also has partners from different countries. Other corpora aim at gathering a large amount of learner data from one specific country and language background (e.g Jinan Chinese Learner Corpus) (Wang *et al*. 2015) and *Corpus do Inglês sem Fronteiras* described in this article and also in Dutra et al. (in press). There are also studies carried out in Brazil with the compilation of small learner corpus, which, unfortunately, in most cases, are of restricted use of the researchers themselves (Alcântara, 2015)[4].

The main objective of this paper is to describe the compilation of a Brazilian university level learner corpus (*CorIsF-Inglês*[5]). Such compilation is based on pre-set parameters that are essential for the feasibility of subcorpora[6] comparison. These parameters are described in the next section. In order to provide a sample of the type of analysis that Corpus Linguistics can facilitate, a partial data analysis is presented based on lexical frequency. Data was extracted from dependent and independent tasks with the purpose of yielding insights to our understanding of Brazilian learners' interlanguage.

2. METHODOLOGY

In this part of the paper we describe *CorIsF-Inglês* according to the following characteristics: participants, data collection, corpus design and data analysis.

2.1 Participants

As mentioned in the introduction, Brazil has a university level *corpus*, Br-ICLE, compiled at several universities in the country, yet, it is restricted to texts produced by English major students, which makes it quite different from the corpus we are describing in this article, namely *CorIsF-Inglês*. This corpus, due to its objective of compiling Brazilian university level student interlanguage, focuses on the collection of texts composed by participants from different college courses. Students who have given us permission to use their texts for the corpus are registered at the English without Borders Program face-to-face courses. The target audience comes mainly from the hard sciences and health courses. Nevertheless, students from arts and humanities have been able to register in some universities, depending on the offering of English courses and the students' interest in taking them[7]. The insertion of their texts in *CorIsF-Inglês* is authorized after they read the consent form and agree with its terms[8]. Although at first all the texts are identified, so that teachers can give their students feedback, as soon as they are sent to the corpus managers, they are given a number. Other major differences between Br-ICLE and *CorIsF-Inglês* lie in mode task variety and in type

[3] <http://www-user.uni-bremen.de/~callies/ALV.htm>.
[4] <http://www.pgletras.uerj.br/linguistica/textos/livro02/LTAA02_a05.pdf>.
[5] *CorIsF-Inglês* is available at <https://sites.google.com/site/corpusisf/home>.
[6] This article data comes from tasks collected at UFMG, but soon other university *subcorpora* will be available to partners' analysis.
[7] Undergraduate or graduate student course is one of the metadata information participants provide. This can be retrieved for research purposes and will be described later on in the section.
[8] A model of the informed consent form is in Appendix A. Each *CorIsF-Inglês* partner university may have a slightly different consent form based on the guidelines of their own university ethics committee.

of data collection. While Br-ICLE concentrated on written interlanguage, *CorIsF-Inglês* has been designed to compile both written and oral interlanguage. Moreover, Br-ICLE collaborators compiled argumentative essays, yet, *CorIsF-Inglês* researchers have aimed at collecting a variety of academic written genres, such as abstracts, summaries and essays. Another difference is that Br-ICLE provides data for cross-sectional studies and *CorIsF-Inglês* design can yield both cross-sectional and longitudinal data.

2.2. Data collection

The data collected for *CorIsF-Inglês* come from tests and activities designed for the *IsF* English courses which show that the primary motivation is to meet students' needs and, consequently, compile a *corpus*. *IsF* teachers have worked together to prepare integrated skills online tests for level A2 (high basic), B1 (intermediate), B2 (high intermediate) and C1 (advanced)[9] so as to prepare their students to take proficiency tests (Dutra, in press). It is clearly evident that most *IsF* audience has the interest to take international proficiency tests, such as TOEFL ITP, TOEFL iBT or IELTS[10] because they have plans to apply for academic scholarships abroad. Despite the fact that most *IsF* courses are not preparatory for proficiency tests, it seems to be reasonable to give students chances to take in-class tests. These tests may help learners develop skills that will enable them to demonstrate their linguistic knowledge even under time constraints. As soon as the idea of preparing online tests was presented, the teachers realized that these tests results would allow them to keep a record of learners' linguistic development for pedagogic and research purposes. Course activities that have also generated texts for *CorIsF-Inglês* are the ones proposed in skill specific courses, such as in academic writing or academic speaking. Text genres produced in these courses are, for instance, summaries and oral presentations.

As mentioned before, the first motivation for online test preparation is pedagogic. When students take a 64-hour course, they can take the test at 3 different points in the term (beginning, middle and end of the course). In addition, students that take part in the *IsF* different level courses for more than one term may have their texts compiled in more than one term[11]. From a pedagogic perspective, teachers have their same level student comparative samples, which allows for the preparation of tailor-made activities to cater for students' needs. Teachers can provide specific feedback to their students and/or adapt course materials based on corpus analysis. Since data and metadata are available for *CorIsF-Inglês* partner institutions, several types of analysis can be done.

[9] The *IsF* course levels are based on the Common European Framework of Reference (CEFR) <http://isf.mec.gov.br/ingles/pt-br/qual-e-meu-nivel-de-proficiencia-em-ingles> and they have been offered from basic level (A2) on.

[10] The acronyms are TOEFL ITP (Test of English as a Foreign Language – Institutional Testing Program); TOEFL iBT (Test of English as a Foreign Language internet-based test) and IELTS (International English Language Testing System).

[11] All research compiled texts do not carry participants' identification. It is the research group responsibility to look for learners' texts produced at different points in time, so as to include them in the longitudinal section of the corpus. The cross-sectional part of the corpus, presented in this article, carries texts that were produced by different participants. In other words, these participants contributed only once to this part of the corpus.

Online test preparation involves the following steps (Dutra et al., in press):

- opening a Gmail account for all the teachers' group to have access to files created in Google Docs;
- taking technical decisions in group:
  - what format the questions should have (multiple choice or open-ended);
  - how to save files in the office computer or in personal computers;
  - which internet resources could be used (e.g. *Youtube*, *TED-Ed*);
  - how to give feedback to students (e.g. automatic feedback using a free online program called *Flubaroo*[12]);
- choosing test themes;
- making small teacher groups according to test themes and CEFR level, so they could prepare the tests;
- sharing activities to receive other teachers' and English Teaching Assistants' (ETAs[13]) feedback;
- making tests available to students through Google Docs;
- sending automatic results to students and teacher;
- sending written or oral texts to the teacher for group and individual feedback;
- sending the results to *Cor-IsF Inglês* for storage and for sharing them with partners.

2.3. *Corpus* design

The design of *CorIsF-Inglês* (Table 1) allows for the compilation of oral and written learner language in a variety of genres. Data comes from timed activities such as tests or online activities (e.g. argumentative essay or opinion response) or may be the result of preparation and/or several drafts as in course activities (e.g. presentations or abstracts) that are process-oriented. Therefore, data can be sorted out depending on research interest. For instance, researchers may analyze verb tense usage in timed and untimed activities so as to depict appropriateness of tense variation.

Table 1: *CorIsF-Inglês* design

Data source	Mode	Genre[14]	Production conditions	Type of task
online test	written	argumentative essay	timed and single draft	independent
		descriptive or comparative essay		integrated
course activities	written	summary; abstract; e-mail; essay	untimed and multiple drafts	integrated
course activities	oral[15]	presentation; debate	untimed and multiple drafts	integrated
online activity	oral	opinion response based on personal experience	timed and single draft	independent
		descriptive or opinion response	timed and single draft	integrated

[12] <http://www.flubaroo.com/>.
[13] The English Teaching Assistant Program is sponsored by CAPES and Fulbright, giving support to *IsF* activities.
[14] The genres listed in Table 1 are examples of genres that have been or can be collected for *CorISF-Inglês*. Other textual genres can be added to the list depending on teachers' suggestions, especially if different courses are prepared as long as they are academic-oriented. As for the oral texts produced through online activities they are responses to a prompt or to texts students listened to or read. Responses are not textual genres and, in fact, they may be interpreted as a function (e.g. giving opinion, making comparisons or presenting a description). Therefore, such tasks require learners to develop strategies to produce an answer that is according to a specific communicative purpose.
[15] *CorIsF-Inglês* oral mode is being designed as technical aspects need to be carefully set before data collection starts.

Data (students' texts) and metadata (information about text genre, text production conditions, participant's age, TOEFL score, course, etc.) are saved in .csv files (comma-separated values), making them organized in spreadsheets that may be easily manipulated by teachers and research partners. Before data is made available to all partners, they are carefully screened so only authorized texts become part of the corpus. Each text receives a reference number and is cleaned (e.g. typos, letter and word repetition are removed). Data is treated with *R*, which is a free software for statistics and graphics that can be widely used in corpus linguistics[16].

One of the corpus characteristics that can be singled out is the type of tasks that the participants were involved with: independent and integrated tasks. Independent tasks require that learners use their world knowledge and personal experiences to produce texts, such as in argumentative essays (written mode) or opinion responses (oral mode). On the other hand, integrated tasks make participants use information presented in written and oral texts or even in infographics or graphs. They are, thus, asked to select and report information, using the criteria of relevance to make comparisons. Lexical-grammatical patterns seem to be influenced by type of task proposed (Biber & Gray, 2013) and such tendency needs to be thoroughly investigated in *CorIsF-Inglês* due to our interest in better understanding Brazilian learners' interlanguage at different acquisition stages.

2.4. Generating and analyzing data

Using the R software, frequency lists with and without stopwords were generated helping the partial analysis of the corpus. Stopwords are words that carry little informational content, e.g. *an, the, on, in,* etc. These lists have been used to create word clouds that show visually the prominence of words in a corpus. At this early stage of our research, parts of the corpus are unbalanced, which means that independent tasks have yielded much more data than integrated tasks. Word frequency lists and word clouds have been used to organize our data for analysis. The most frequent grammatical categories were identified and a connection with their frequency in specific genres was correlated.

3. *CORISF-INGLÊS* PARTIAL ANALYSIS

The data presented in this section illustrates what type of analysis can be carried out based on a learner corpus. There is an array of possible research foci and we emphasize that this sample analysis can be greatly improved once the corpus grows and it is balanced. A lexical analysis will be presented considering the two parts of the corpus: data from independent task students' production and data from integrated tasks. Consequently, we discuss the grammatical categories that tend to emerge from these two types of tasks.

The independent task data is comprised of 104,437 words; therefore, it is the majority of *CorIsF-Inglês* data that up to this date has 130,999 words. This data shows that most online tests prepared by our *IsF* group included

[16] <http://www.r-project.org/>.

this type of writing task[17]. Besides generating a frequency list, we wanted to have a visual idea of which words tend to be more produced by learners in independent written tasks. Therefore, word clouds were generated to provide a picture of the most prominent words in this part of the corpus.

Figure 1 presents the word cloud for independent task data without stopwords, revealing that the spotlight is given to nouns and auxiliary modal verbs (*people*, *person, can, will*) as well as to adjectives (*good* and *important*) and verbs (*think* and *like*). In a very similar level of frequency there is also an adverb (*first*) and the word *one* which can be classified in different ways, according to its function in the sentence, for instance as a determiner or a pronoun[18]. It calls our attention that in the next level of frequency there are many nouns (e.g. *religion, language, students, impression, water, school* etc.) probably chosen due to task topics and prompts. Verbs are the second most prominent grammatical category in the word cloud (Figure 1) which include the main verbs *think, like, know want, learn* and *make* as well as the auxiliary modal verbs *can* and *will*. Most of these verbs are mental state verbs (*think, know, want* and *learn*), making evident that the expression of participants' ideas and thoughts are central in independent tasks which require the production of position or argumentative essays (see Appendix C for an example). In such tasks writers are supposed to present ideas and to convince the readers of an issue.

Figure 1: Independent task word cloud without stopwords

The integrated task wordcloud (Figure 2) was generated from the 26,562 words, excluding the stopwords, which were compiled from the *CorIsF-Inglês* integrated task section. The much lower number of words in this corpus part, as compared to the independent task part, as already mentioned, is largely due to the fact that most online tests included independent tasks rather than the integrated ones. There is a prevalence of nouns in the word cloud

[17] It is our plan to have a wide variety of online tests available to be used, which should also reflect on a more balanced assortment of written tasks.

[18] Another level of analysis would have to be done in order to distinguish the function of such words in each sentence.

in Figure 2, such as *coffee*[19], *people* and *day*. The next most frequent group of words includes mainly nouns as well (e.g. *year, Americans, cups, caffeine, men* and *women*). The verbs that are prominent are *can* (auxiliary modal verb), *drink*, *divorced* and *consumed*. The last two forms may also have been used in the tasks as adjectives. There were two test integrated task themes that seem to have attracted most *IsF* teachers and the word cloud clearly represents this preference (Coffee theme for A2 level and Love, Marriage and Divorce for B1). Nouns are, then, the most frequent grammatical items in the integrated task word cloud. This is not a surprise to us as integrated tasks require learners to describe and report information provided in infographics or charts, making comparisons if appropriate.

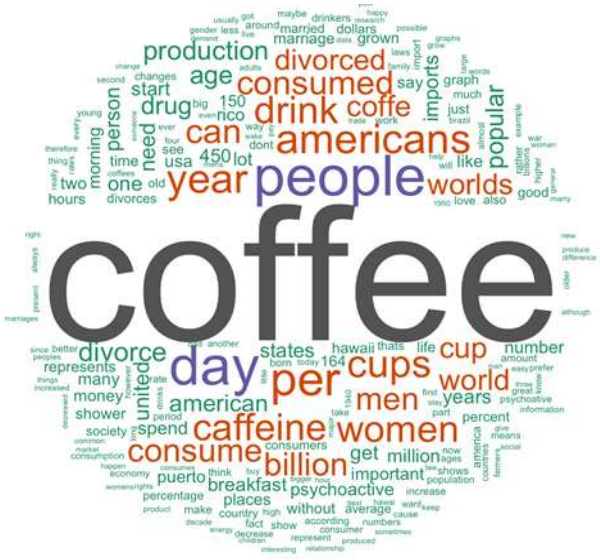

Figure 2: Integrated task word cloud without stopwords

Comparing the results from word frequency analysis of independent and integrated tasks, some differences arise. The type of task is likely to have influenced the frequency of nouns, verbs and adjectives learners chose to use. While independent tasks tend to favor the high occurrence of nouns, mental state verbs and some adjectives, integrated tasks create the environment for prominent use of nouns. In the former case, arguments prevail and writers need to convince the audience of their position. Mental state verbs, for example, come in handy in argument construction. In the latter type of task, description is valued and verbs, although present in the texts, do not need to be the same. The variation of verb types, then, allows nouns to appear as the most frequent category in integrated writing tasks. Some nouns are clearly at a high frequency level because they are directly related to the task topic (e.g. *coffee*, *divorce*).

[19] A misspelled version of the word (*coffe*) is also prominent in the independent task word cloud. A corpus treatment needs to be done to eliminate such data distortions.

4. CONCLUSION

In this article we presented *Cor-IsF Inglês* design and explained how pedagogical motivations can be linked to research interests. A detailed design allows teachers and research group members to profit from data collection and systematize both group and individual feedback besides creating a corpus that can be widely investigated. The *IsF* community has a great chance to develop an array of studies that may focus either on written or on oral interlanguage or on the comparison of both modes.

It is also relevant to further investigate the results generated from independent and integrated tasks to discuss task design and, furthermore, research how different level students perform in such tasks. A deeper analysis of the corpus, especially when it reaches 200,000 words at each proficiency level, should shed light into a better understanding of interlanguage. From a pedagogic perspective, *IsF* teachers, who are also task planners, should look into corpus data to assist their learners to better express themselves in specific genres.

REFERENCES

Alcântara, Christiane Fontinha de Alcântara. 2015. Intensificação sob a ótica da Linguística de *Corpus*: Uma investigação sobre o inglês oral de aprendizes brasileiros. Retrieved from <http://www.pgletras.uerj.br/linguistica/textos/livro02/LTAA02_a05.pdf> on June 18th, 2015.

Berber Sardinha, Antonio Paulo. 2001. O *corpus* de aprendiz Br-ICLE. *Intercâmbio* 10, p. 227-239.

Biber Douglas; Gray, Bethany. 2013. Discourse Characteristics of Writing and Speaking Task Types on the *TOEFL iBT*® Test: A Lexico-Grammatical Analysis. *TOEFL iBT*® *Research Report* 2013. Acessado em <http://www.ets.org/Media/Research/pdf/RR-13-04.pdf>.

Chomsky, Noan. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Dutra, Deise Prina. In press. Formação de professores: colaboração sem fronteiras. In Dalacorte, Maria Cristina Ferreira (org.).

Dutra, Deise Prina & Berber Sardinha, Tony. 2013. Referential expressions in English learner argumentative writing. In Granger, Sylviane, Gilquin, Gaëtanelle, & Meunier, Fanny (eds.). *Twenty Years of Learner Corpus Research: Looking back, Moving ahead*. *Corpora and Language in Use* – Proceedings 1. Louvain-la-Neuve: Presses universitaires de Louvain, p. 117-127.

Dutra, Deise Prina, Gomide, Andressa, Oliva, Katherine, & Guedes, Annallena. In press. *Corpus* de aprendizes do Inglês sem Fronteiras: caminhos para compreender a interlíngua de alunos universitários brasileiros. In Sarmento, Simone, Abreu-e-Lima, Denise Martins de, & Moraes Filhos, Waldenor Barros. *Do Inglês sem Fronteiras ao Idiomas sem Fronteiras:* na construção de uma política linguística para a internacionalização.

Granger, Sylviane (ed.). 1998. *Learner English on Computer*. London & New York: Addison Wesley Longman.

Granger, Sylviane, Dagneaux, Estelle, Meunier, Fanny, & Paquot, Magali. 2009. *The International Corpus of Learner English.* Version 2 Handbook and CD-ROM. Louvain-la-Neuve: Presses universitaires de Louvain.

Granger, Sylviane. 2015. Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research* 1 (1), p. 7-24.

Leech, Geoffrey. 1998. Preface. In Granger, S. (ed.), *Learner English on Computer*. London & New York: Addison Wesley Longman.

Mello, Heliana, Avila, Luciana; Neder Neto, Tufi, & Orfano, Barbara. 2012. LINDSEI-BR: an oral English interlanguage *corpus*. In VII GSCP International Conference: Speech and *Corpora*, 2013, Belo Horizonte. *Proceedings* of the VII GSCP International Conference: Speech and *Corpora*. Florença, Itália: Firenze University Press, p. 85-86.

MacWhinney, Brain. 1987. The competition model. In MacWhinney, Brain (ed.). *Mechanisms of language acquisition*. Hillsdale NJ: Erlbaum.

Schachter, Jacquelyn. 1988. Second language acquisition and its relationship to universal grammar. *Applied Linguistics* 9 (2), p. 219-235.

Selinker, Larry. 1972. Interlanguage. *International Review of Applied Linguistics* 10, p. 209-231.

Shepherd, Tânia. 2009. *Corpora* de aprendiz de lingua estrangeira: um estudo de n-gramas. *Veredas* 2, p. 100-116.

Wang, Maolin, Malmasi, Shervin, & Huang, Mingxuan. 2015. The Jinan Chinese Learner *Corpus*. *Proceedings of the Tenth Workshop on Innovative Use of NLP for Building Educational Applications*. Denver: Association for Computational Linguistics.

White, Lydia. 1989. *Universal grammar and second language acquisition*. Philadelphia: John Benjamins.

APPENDIX A

CARTA DE CONSENTIMENTO LIVRE E ESCLARECIDO:

Para os participantes

Caro(a) Senhor(a)

A coordenação do Programa Inglês sem Fronteiras/UFMG conduz pesquisas que visam estudar o desenvolvimento das habilidades de leitura, de escrita, de audição e de fala de aprendizes de língua inglesa para fins acadêmicos. Cada projeto de pesquisa está devidamente autorizado pela Câmara de Pesquisa da Faculdade de Letras da UFMG.

A fim de que os projetos possam ser desenvolvidos, é necessária a sua autorização, vez que as pesquisas constarão da coleta das suas redações produzidas enquanto aluno do curso. A sua participação nesta pesquisa é voluntária e não determinará qualquer risco nem trará desconfortos. Além disso, sua participação é importante para o aumento do conhecimento a respeito dos processos de aquisição e desenvolvimento das quatro habilidades supracitadas por alunos universitários brasileiros, podendo beneficiar outros alunos futuramente na melhoria do ensino de língua inglesa no nível superior.

Informamos que o/a Sr(a). tem a garantia de acesso, em qualquer etapa dos estudos, sobre qualquer esclarecimento de eventuais dúvidas. Se tiver alguma consideração ou dúvida sobre a ética da pesquisa, entre em contato com o Comitê de Ética em Pesquisa (CoEP) da Universidade Federal de Minas Gerais, situado na Av. Antônio Carlos, 6627. Unidade Administrativa II - 2° andar - Campus Pampulha, telefone 3409-4592 / 3409-4027.

Também é garantida a liberdade da retirada de **consentimento** a qualquer momento e deixar de participar do estudo.

Fica também garantido que as informações obtidas serão analisadas em conjunto com as de outras pessoas, não sendo divulgada a identificação de nenhum dos participantes.

O/A Sr(a). tem o direito de ser mantido atualizado sobre os resultados parciais das pesquisas e, caso seja solicitado, todas as informações que solicitar lhe serão fornecidas.

Não existirão despesas ou compensações pessoais para o participante em qualquer fase dos estudos. Também não há compensação financeira relacionada à sua participação.

Os participantes das pesquisas comprometem-se a utilizar os dados coletados somente para pesquisa, e os resultados serão veiculados através de artigos científicos, em revistas especializadas e/ou em encontros científicos e congressos, sem nunca tornar possível a sua identificação.

Abaixo se encontra o Termo de **Consentimento Livre e Esclarecido**, para ser assinado caso não tenha ficado qualquer dúvida.

Deise Prina Dutra – Coordenadora Geral do IsF/UFMG

Ana Larissa Adorno Marciotto Oliveira – Coordenadora Pedagógica do IsF/UFMG

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Acredito ter sido suficientemente informado a respeito dos estudos conduzidos pela coordenação do Programa Inglês sem Fronteiras/XXXX. Ficaram claros para mim quais são os propósitos dos estudos, os procedimentos a serem realizados, as garantias de confidencialidade e de esclarecimentos permanentes. Ficou claro, também, que a minha participação é isenta de despesas e que tenho garantia do acesso aos resultados e de esclarecer minhas dúvidas a qualquer tempo. Concordo voluntariamente em participar e estou ciente de que poderei retirar o meu consentimento a qualquer momento sem penalidade ou prejuízo ou perda de qualquer benefício que eu possa ter adquirido.

○ Concordo
○ Discordo

APPENDIX B

INTEGRATED WRITING TASK

Coffee (A1 - A2)

You should spend about 20 minutes on this task and write at least 150 words.

The infographic below presents some information about coffee. Organise the information by selecting and reporting the main features, and make comparisons where relevant.

Retrieved from <http://www.designinfographics.com/food-infographics/a-coffee-crazed-america>.

APPENDIX C

INDEPENDENT WRITING TASK

Languages (B2)

Read the question below. Give yourself 30 minutes to plan, write, and revise your essay. Typically, an effective response will contain a minimum of 300 words.

- Do you agree or disagree with the following statement? Children should begin learning a foreign language as soon as they start school. Use specific reasons and examples to support your position.

Submitted: 08/07/2015
Accepted: 05/10/2015