

# O CORPUS LILE: ABSTRACTS DAS ÁREAS DE LINGUÍSTICA E LITERATURA

Simone Sarmento<sup>1</sup>  
Bruno Scortegagna<sup>2</sup>  
Larissa Goulart<sup>3</sup>

## 1. Introdução

A presente pesquisa, relatada parcialmente neste artigo, iniciou-se a partir da percepção compartilhada entre professores de língua inglesa da Universidade Federal do Rio Grande do Sul (UFRGS) de que há uma carência de materiais didáticos baseados em textos autênticos para o ensino de Inglês Acadêmico. Dessa forma, nos propusemos a compilar um *corpus* de gêneros acadêmicos das áreas de Literatura e Linguística, de forma a subsidiar a análise linguística e a elaboração de projetos e tarefas pedagógicas com foco no acadêmico de letras, futuro professor de línguas e/ou tradutor.

A compilação do *corpus* LILE (*Corpus* de Gêneros Acadêmicos de Linguística e Literatura) se dará em etapas. A primeira etapa, que consistiu na compilação de um *corpus* de *Abstracts*, já está concluída. A segunda fase da compilação que ainda está por vir será de artigos acadêmicos também da área de Letras. Portanto, nesse artigo iremos explorar um dos *subcorpora* já compilado e também demonstraremos uma tarefa voltada para o ensino desse gênero (*abstract*), em inglês.

Um *abstract*<sup>4</sup> é um breve resumo de um artigo de pesquisa, tese, revisão, conferência, ou qualquer análise aprofundada sobre um determinado assunto ou disciplina, e é frequentemente usado para ajudar o leitor a tomar ciência rapidamente do propósito do artigo. Neste estudo focalizamos apenas os *abstracts* presentes em textos acadêmicos (artigos científicos, trabalhos de conclusão de curso, dissertações e testes), que aparecem no início de um manuscrito, atuando como o ponto de entrada para os textos acadêmicos. Nesse sentido, não serão abordados, por exemplo, os *abstracts* utilizados como a base para a seleção de propostas para apresentação em uma conferência acadêmica.

Decidimos iniciar a compilação com o gênero *abstract* primeiramente por esse apresentar acesso facilitado. *Abstracts*, em geral, são de domínio público, mesmo em revistas em que o acesso se faz de forma privada<sup>5</sup>. Em segundo lugar, como já mencionado, os *abstracts* devem ser a síntese do texto, que atraem a atenção do leitor. Segundo Swales e Feak (1994)<sup>6</sup>:

---

1 Professora Adjunta do Instituto de Letras-PPGLetras/UFRGS

2 Bolsista de IC – Acadêmico do Instituto de Letras/UFRGS

3 Bolsista de IC – Acadêmico do Instituto de Letras/UFRGS

<sup>4</sup> Optamos por utilizar *abstracts*, no lugar de *resumo*, que seria o correspondente em língua portuguesa, uma vez que o último é polissêmico, podendo referir-se ao texto inicial tratado aqui neste trabalho ou, ainda, ao fato de coletar e redigir as ideias principais de um texto.

<sup>5</sup> Alguns periódicos, principalmente os internacionais, só abrem acesso mediante o pagamento por assinaturas anuais ou de artigos individuais. Entretanto, os *abstracts* estão sempre disponíveis.

<sup>6</sup> Tradução feita pelos autores. Versão original: “The research world is facing “an information explosion” with several million research papers being published each year. There are also continual announcements of new journals being launched, either online or in hard copy or via both types of delivery. Many researchers have therefore to be highly selective in their reading, often focusing on skimming abstracts and key words. Journal article (JA) abstracts have thus become an increasingly important part-genre.”

“O mundo das pesquisas está enfrentando uma “explosão de informações” com milhões de artigos de pesquisas sendo publicados a cada ano. Há também anúncios contínuos de novas revistas sendo lançadas, de forma online ou impressas, ou das duas formas. Muitos pesquisadores precisam, portanto, ser altamente seletivos em suas leituras, enfocando, geralmente, a leitura rápida dos resumos e palavras-chaves. O *abstract* de periódicos se torna assim cada vez mais um subgênero importante.”

Em terceiro lugar, concordando com Dayrell (2011) ao afirmar que, no Brasil, “a necessidade de produção de resumos científicos em inglês é evidente, pois são obrigatoriamente parte de dissertações e teses e, geralmente, incluídos em artigos científicos em português, seja para periódicos acadêmicos ou anais de congressos” (p.139). Utilizando a metodologia da Linguística de *Corpus*, a autora pesquisou as diferenças potenciais entre *abstracts* em inglês escritos por alunos de pós-graduação brasileiros e aqueles extraídos de artigos publicados e redigidos por escritores mais proficientes. Nossa pesquisa está em consonância com Dayrell, na medida em que compilamos um *corpus* de *abstracts* produzidos por escritores com diferentes graus de proficiência na escrita acadêmica, ou seja, desde alunos de graduação (TCCs), até pesquisadores experientes e de renome internacional (Revistas *Qualis A* - internacionais), passando também por alunos de pós-graduação<sup>7</sup>. Essa característica possibilita investigar até que ponto as produções dos alunos se aproximam ou se distanciam daquelas publicadas em revistas de grande circulação e que são aceitas entre a comunidade acadêmica, consideradas, aqui, como o padrão de excelência a ser atingido. Entretanto, enquanto Dayrell investigou *abstracts* de áreas científicas compreendendo a física, a computação e a farmácia, a presente pesquisa foca exclusivamente em textos da área de Letras, área aparentemente pouco contemplada nas pesquisas baseadas em *corpus*. Outro ponto a ser comentado é o fato de que, no desenrolar deste projeto, não faremos referência à recorrente dicotomia falante nativo e falante não nativo. Acreditamos que neste tipo de gênero é difícil precisar o verdadeiro redator do texto, pois muitos alunos de graduação, de pós-graduação ou até mesmo pesquisadores experientes, pedem auxílio para redigir um *abstract* em Inglês. A diferença principal entre *abstracts* de TCCs/Teses/Dissertações, ou seja, produzidos por alunos e aqueles presentes nas publicações *Qualis A*, produzidos por pesquisadores, reside precisamente no diferente grau de revisão que recebem. Enquanto trabalhos acadêmicos não publicados não possuem uma padronização ou exigência de revisão, periódicos de qualidade reconhecida geralmente passam por vários revisores/pareceristas antes de sua publicação, o que confere um caráter de maior aceitação do texto.

Esta pesquisa poderá contribuir tanto para o ensino de língua adicional<sup>8</sup> quanto para pesquisas de cunho mais analítico. A pesquisa e seus resultados irão beneficiar primeiramente o próprio Instituto de Letras (IL) da Universidade em que é desenvolvida, e, em segunda instância, a comunidade acadêmica como um todo. O *corpus* de estudo poderá ser utilizado para a elaboração de tarefas por outros professores, pois o material compilado está armazenado na base de dados do projeto

---

<sup>7</sup> O *corpus* será descrito em detalhes na seção 3.

<sup>8</sup> Schlatter e Garcez (2009) preferem o termo “língua adicional” no lugar de língua estrangeira devido ao acréscimo que as línguas inglesa e espanhola trazem para quem se ocupa delas e por serem línguas que, muitas vezes, estão a serviço da comunicação entre pessoas de diversas culturas e nacionalidades, sendo difícil definir claramente quem são os nativos ou os estrangeiros. Os autores consideram que, de certa forma, “essas línguas fazem parte dos recursos necessários para a cidadania contemporânea” (op cit, p. 128).

TERMISUL<sup>9</sup>, estando disponível para diversos tipos de consulta e para ser utilizado com variadas ferramentas (tal como concordanciador e gerador de N-gramas).

## 2. Metodologia e desenvolvimento

A Linguística de *Corpus* foi utilizada como metodologia na crença de que, para o objetivo proposto, seja um método bastante eficaz para conhecer a realidade e peculiaridades dos textos estudados. Conforme Leech (1992, p. 111), “a observação contribui mais para a teoria do que a teoria contribui para a observação”. É preciso ressaltar, neste momento, que cada pesquisa desenvolvida a partir de uma metodologia baseada em *corpus* é única. Sendo assim, os procedimentos adotados nesse tipo de pesquisa são norteados pelos dados e pelas necessidades impostas pela pesquisa no decorrer do processo de análise.

Há alguns aspectos da Linguística de *Corpus* que merecem ser destacados:

- A Linguística de *Corpus* vê a língua como sendo um sistema probabilístico, ou seja, embora muitas vezes combinações e características linguísticas sejam possíveis, nem todas são prováveis de ocorrer (STUBBS, 2001). Os *corpora* são usados para gerar conhecimento empírico sobre uma língua, que pode suplementar, ou muitas vezes suplantar, informações provenientes de fontes de referência e introspecção (LEECH, 1991, 1992);
- O trabalho é empírico, estando calcado na análise de exemplos e padrões reais de uso em textos naturais;
- A quantidade de material analisado é sempre robusta, deve ser sempre considerável para os propósitos que se propõe;
- A Linguística de *Corpus* utiliza ferramentas digitais para análise: automáticas, semi-automáticas e manuais.
- A investigação depende de técnicas de análise quantitativa e qualitativa.

O principal objetivo da Linguística de *Corpus* é estudar os padrões e as variações sistemáticas que a língua apresenta. Conforme Sarmiento (2008) “a variação sistemática, ou seja, a recorrência de traços linguísticos (colocação, coligação, padrão sintático, entre outros) indica que a linguagem é padronizada (*patterned*) e motivada por diversos fatores além das necessidades comunicativas”. Portanto, um dos nossos objetivos é mostrar como os diferentes tipos *abstracts*, de alguma maneira, são padronizados, e como essa característica pode contribuir para o ensino desse gênero.

Para realizar a análise dos *abstracts*, foi realizada uma comparação do *subcorpus* de aprendizes com o *subcorpus* de usuários mais experientes<sup>10</sup>. Acreditamos que o tipo de investigação aqui proposta poderá detectar os tipos de problemas que os aprendizes encontram quando escrevem um texto acadêmico. Para o desenvolvimento de tarefas pedagógicas relevantes e fidedignas foi necessário analisar os diversos *subcorpora* que

---

<sup>9</sup> O projeto aqui relatado é parte integrante de um Objeto de Aprendizagem (AO) que contemplou também as línguas portuguesa, espanhola e francesa. A compilação do corpus de língua espanhola teve a supervisão da Professora Cleci Bevilacqua, e o de língua francesa foi supervisionado pela professora Sandra Loguércio. Além dos corpora, o Objeto de Aprendizagem apresenta tarefas que visam auxiliar os alunos do curso de Letras, assim como profissionais interessados na linguagem acadêmica, a familiarizarem-se com o gênero abstract. Os corpora e as tarefas podem ser encontrados no link <http://www6.ufrgs.br/termisul/oa3/>.

<sup>10</sup> *Corpora* de aprendizes são geralmente entendidos como uma base de dados de textos escritos por falantes menos proficientes de uma língua (neste trabalho não usaremos o tradicional paradigma variante nativa *versus* variante não nativa por motivos já explicitados anteriormente).

serão citados na descrição da compilação do *corpus*, com o intuito de verificar os aspectos linguísticos mais divergentes entre os usuários mais e menos experientes (GRANGER et al, 2007).

## 2.1 Compilação do *corpus*

O *corpus* de *abstracts* compreende a produção de textos publicados entre 2008 a 2011. Essa faixa cronológica de textos representa uma linguagem atual e usual, ponto que é de extrema importância para a produção de material didático, pois assim as tarefas/unidades didáticas criadas terão a possibilidade de refletir uma visão contemporânea de uso da língua. A coleta dos *abstracts* da área de Letras da UFRGS se deu a partir de uma base de dados mantida pela universidade, conhecida como LUME, onde se encontram todos os TCCs, Teses e Dissertações produzidos na universidade nos últimos anos. A seleção das revistas se deu a partir da classificação das mesmas, pois, forma selecionadas apenas revistas nacionais e internacionais que tiveram o conceito Qualis A no triênio 2009-2011. Foram selecionadas duas revistas nacionais de cada área, Literatura e Linguística. Esses periódicos tiveram seus *abstracts* compilados também a partir do ano de 2008 até o ano de 2011, quando disponíveis nos seus sites. A seleção das revistas internacionais transcorreu da mesma forma.

O *corpus* de *abstracts* de Língua Inglesa está dividido em oito *subcorpora*, totalizando 700 textos: (1) TCCs: 44 de linguística e 42 de literatura; (2) Teses e Dissertações: 71 de linguística e 79 de literatura; (3) Revistas Nacionais: 120 de linguística e 100 de literatura; e (4) Revistas Internacionais: 116 de linguística e 128 de literatura. Para os grupos (1) e (2), compilamos a totalidade de trabalhos presentes no LUME, publicados entre 2008 e 2011. A partir da quantidade de textos nesses *subcorpora*, decidimos a respeito da quantidade de periódicos a ser incluído nos outros dois grupos, de forma a manter um certo equilíbrio entre o número de *abstracts* em cada um dos *subcorpora*.

As tabelas a seguir mostram, de forma mais clara, os números do *corpus* de *abstracts*:

**Tabela 1** – Número de textos no *corpus* LILE:

	LINGUÍSTICA	LITERATURA
TCCs	44	42
TESES E DISSERTAÇÕES	71	79
REVISTAS NACIONAIS	120	100
REVISTAS INTERNACIONAIS	116	128
TOTAL DE TEXTOS EM INGLÊS	700	

**Tabela 2** - Número de itens no *corpus* LILE

	LINGUÍSTICA	LITERATURA
TCCs	9.075	5.946
TESES E DISSERTAÇÕES	21.067	19.487
REVISTAS NACIONAIS	16.018	8.017
REVISTAS INTERNACIONAIS	19.488	23.041
TOTAL DE ITENS EM INGLÊS	122.139	

## 3. Análise dos *subcorpora*

Apresentamos a seguir a comparação e análise de dois *subcorpora*: TCCs – Linguística e Revistas Internacionais – Linguística. Essa etapa é fundamental para o desenvolvimento de uma tarefa pedagógica. O programa que utilizamos para a análise dos textos foi o WordSmith Tools V.5 (SCOTT, 2009).

Primeiramente, geramos uma lista de palavras mais frequentes dos dois *subcorpora*. Isso permite a visualização das palavras mais e menos recorrentes (VIANA, 2011). Esse procedimento revela vários aspectos sobre os textos incluídos no *corpus* e gera hipóteses investigativas que precisarão ser complementadas com análises mais detalhadas.

**Tabela 2** – Lista de palavras dos *subcorpora* Revistas Internacionais e TCCs

	LINGUÍSTICA	
<i>Corpus</i> de	Revistas Internacionais	TCCs
1	THE	THE
2	OF	OF
3	AND	IN
4	IN	TO
5	TO	AND
6	A	A
7	LANGUAGE	THAT
8	THAT	THIS
9	AS	IS
10	ON	LANGUAGE
11	FOR	AS
12	THIS	ON
13	IS	FOR
14	ARE	BY
15	WITH	TEACHING
16	ENGLISH	STUDENTS
17	BY	WHICH
18	IT	AN
19	WHICH	FROM
20	AN	WITH

Como podemos ver na tabela acima, a maioria das palavras é gramatical, tal como *the, of, etc.*, com baixo valor semântico, o que já era esperado. Utilizamos a ferramenta *concord* para verificar as linhas de concordância da palavra *language*, pois é a palavra lexical mais frequente nos dois *subcorpora*. A recorrência de *language* deve-se ao teor temático da maioria dos *abstracts*, que gira em torno desse assunto. A seguir, nas Tabelas 3 e 4, mostramos os *clusters* de três palavras mais frequentes, com uma frequência mínima de cinco ocorrências, gerados a partir da palavra nódulo *language*.

**Tabela 3** - Clusters gerados no *subcorpus* Revistas Internacionais/Linguística/Inglês

N	Cluster	Freq.
1	SECOND LANGUAGE ACQUISITION	13
2	THE LANGUAGE OF	7
3	A SECOND LANGUAGE	7
4	SECOND LANGUAGE L2	7
5	OF THE LANGUAGE	6
6	LANGUAGE LEARNING AND	5
7	ON SECOND LANGUAGE	5
8	AS THE LANGUAGE	5
9	AND SECOND LANGUAGE	5
10	A FOREIGN LANGUAGE	5
11	LANGUAGE IDEOLOGIES THAT	5
12	IN SECOND LANGUAGE	5
13	IN LANGUAGE POLICY	5

**Tabela 4** - Clusters gerados no *subcorpus*<sup>11</sup> TCCs/Linguística/Inglês

N	Cluster	Freq.
1	AN ADDITIONAL LANGUAGE	8
2	LANGUAGE IN THE	7
3	A FOREIGN LANGUAGE	5

<sup>11</sup> O número de *clusters* gerado a partir do *subcorpus* de TCC é menor devido ao tamanho do mesmo, que é constituído de somente 44 textos e 9075 itens, ao passo que o *subcorpus* de Revistas Internacionais é composto de 116 textos e 19488 itens.

Nota-se que no *subcorpus* de TCCs, *an additional language* é o *cluster* mais frequente, mas não apareceu entre os *clusters* de revistas internacionais, que apresentou *second language acquisition* como o mais recorrente. A seguir, algumas linhas de concordância a fim de verificar o cotexto<sup>12</sup> dos principais *clusters*.

**Tabela 5** - Linhas de concordância do *corpus* de RI - *Second language acquisition*

N	Concordance
2	us our comparison of first and <b>second language acquisition</b> a
6	e discussion of age effects in <b>second language acquisition</b> .
7	ion of developmental issues in <b>second language acquisition</b>
9	debated topics in the field of <b>Second Language Acquisition</b> .
10	explanation of age effects on <b>second language acquisition</b> .
13	oken frequencies determine the <b>second language acquisition</b>
15	t refer in any explicit way to <b>second language acquisition</b>

Observando os colocados que aparecem anteriores ao *cluster* “*second language acquisition*” percebemos que esses carregam as ideias do que será analisado no trabalho completo.

**Tabela 6** - Linhas de concordância do *corpus* de RI – *the language of*

N	
1	ols, which have been utilizing <b>the language of</b> community, Ma
2	maths. This meant a change in <b>the language of</b> education for
4	ained: model A with Spanish as <b>the language of</b> instruction a
5	ion and model D with Basque as <b>the language of</b> instruction a
6	used stereotypes on English as <b>the language of</b> science are a

Nessas linhas de concordância, como o *cluster* já sugere, se torna evidente que o uso de “*the language of*” em geral vem acompanhado da referência a um uso especializado da língua. Essas linhas serão relevantes na elaboração da tarefa que será apresentada.

**Tabela 7** - Linhas de concordância do *corpus* de RI – *English as a*

N	

<sup>12</sup> “Cotexto” indica o que está imediatamente à esquerda à direita da palavra de busca enquanto “contexto” pode também referir-se a elementos extra-textuais, como condições de produção, tipos de interlocutores, entre outros fatores.

2	f EIL or de-Anglo-Americanized <b>English as a</b> means of express
3	genized variety of English and <b>English as a</b> lingua franca in
4	have increasingly constructed <b>English as a</b> branded commodit
6	ure of speaking proficiency in <b>English as a</b> second language
7	roup) enrolled in introductory <b>English as a</b> second language
8	articipants were college-level <b>English as a</b> second language

Nos exemplos acima, “*English as a*” faz referência explícita a uma variedade da língua inglesa.

Na análise das linhas de concordâncias dos três primeiros *clusters* de Revistas Internacionais nota-se a recorrência de *Language* e *English*. O primeiro *cluster* “*second language acquisition*” se torna muito interessante quando comparado com o primeiro dos TCCs “*an additional language*” que parece ser recorrente apenas no subcorpus de TCCs.

**Tabela 8** – Linhas de Concordância do *corpus* de TCCs - *An additional language*

N	
1	ontribute to the discussion of <b>an additional language</b> teaching
2	ided into five parts: learning <b>an additional language</b> in the
3	tion of motivation in learning <b>an additional language</b> in the
4	student's motivation to learn <b>an additional language</b> , as we
5	hods of teaching Portuguese as <b>an Additional Language</b> , in Br
6	to the Portuguese teaching as <b>an additional language</b> . This

Se observarmos o uso de *an additional language*, percebemos uma estreita ligação com o verbo “*learn*”, principalmente na forma “*learning*”. Nota-se também a presença de “*teaching*”, “*motivation*” e “*Portuguese*” como colocados.

**Tabela 9** – Linhas de concordância do *corpus* de TCCs – *Language in the*

1	parts: learning an additional <b>language in the</b> language use
2	tion in learning an additional <b>language in the</b> basic public
3	normal language/pathological <b>language in the</b> studies about

*Language in the* também está relacionado ao termo Língua Adicional usado em português. Podemos observar que em duas ocorrências “*language in the*” está associado a “*an additional language*”.

**Tabela 10** – Linhas de concordância do *corpus* de TCCs – a Foreign Language

1	in a classroom of English as a <b>Foreign Language</b> (EFL). The
2	for the teaching of German as a <b>foreign language</b> in schools
3	d to the lack of motivation in a <b>foreign language</b> classroom.
4	sider it the best way to learn a <b>foreign language</b> . It was all
5	tare entering in contact with a <b>foreign language</b> for the first

Essa breve análise servirá como auxílio na elaboração de tarefas pedagógicas voltadas à leitura e redação de *abstracts*. Na próxima seção, iremos abordar essa questão mais detalhadamente.

#### 4. Tarefa pedagógica

Conforme Sardinha (2011), a Linguística de *Corpus* serve como uma ferramenta para professores que se proponham a elaborar uma tarefa que trate dos gêneros textuais e de suas peculiaridades. O gênero *abstract* não se constitui apenas de elementos gramaticais, por isso é preciso ter em mente as suas características particulares, como sua extensão, as escolhas lexicais adequadas, e outros traços que se tornam evidentes através de um *corpus*.

À luz da teoria dos gêneros do discurso, segundo Bakhtin (2006), a língua se realiza socialmente e, portanto, faz parte das atividades humanas. Dentro dessas atividades situa-se a escrita acadêmica, e, no caso deste trabalho, mais especificamente a leitura e redação de *abstracts*. Esses, muitas vezes, como apontado por Swales (1990) e Dayrell (2011), constituem a única parte de um trabalho acadêmico escrita na língua inglesa, mesmo quando a publicação é redigida em outra língua, no nosso caso, a língua portuguesa.

A nossa proposta, nesse sentido, é organizar uma sequência didática que poderá auxiliar o aluno a interpretar e produzir um *abstract* diretamente na língua alvo sem que o aluno precise recorrer a traduções.

A validade da Linguística de *Corpus* como ferramenta para o ensino já foi discutida por vários autores no livro *Corpora no Ensino de Línguas Estrangeiras* (VIANA e TAGNIN, 2011); por O’Keefe, McCarthy e Carter (2007); por McEnery & Wilson (1996:104) entre outros. Os autores sugerem que é importante usar a Linguística de *Corpus* para compor tarefas desde o início da aprendizagem, pois, dessa forma, expomos os alunos à linguagem autêntica, principalmente quando o objetivo do ensino é fazer com que o aluno participe de interações sociais e esteja apto para fazer coisas com a língua no mundo real. Conforme Sardinha (2011, pp.304)

saber uma língua implica conhecer como dizer e escrever segundo as convenções de variedades específicas da língua (um gênero ou registro específico em um contexto determinado); para isso, é preciso conhecer a lexicogramática das escolhas necessárias e desejadas para aquela situação específica. Para usar a lexicogramática com eficiência, é necessário conhecer as probabilidades daquelas escolhas, isto é, as frequências dos elementos, suas combinatórias e as frequências destes.

Em muitos materiais didáticos faz-se o uso de textos não autênticos, ou seja, textos e exemplos elaborados unicamente para a composição das tarefas pedagógicas. Essa prática pode não possibilitar ao aluno ter contato com as características de cada gênero, abrindo lacunas na aprendizagem e tornando a comunicação uma atividade não natural.

#### 4.1 A tarefa

A unidade didática que será apresentada a seguir foi elaborada de forma a suprir as necessidades de alunos de nível intermediário/avançado do curso de Letras de Inglês. Para elaborar a tarefa, utilizamos o programa Reading Class Builder<sup>13</sup> como base, mas fizemos algumas alterações e acrescentamos alguns exercícios para adequar a tarefa aos nossos objetivos<sup>14</sup>.

#### BEFORE READING THE TEXT

1) These are some important words from the texts you are going to read:

Language - English - Study - Research - Students - Teaching - Analysis - Paper - Results - Linguistic - Teachers - Article - Foreign - Use - Based – Process
--

a) What kind of text do you think it is?

b) Can you predict the topic?

#### ANSWER THE FOLLOWING QUESTIONS BASED ON THE TEXT BELOW

2) What kind of text is this? Compare with your first answer.

3) What is the aim of this type of text?

4) What is the general idea of the text?

5) Who could be interested in reading this text?

---

<sup>13</sup> O Reading Class Builder é um software para preparação semi-automática de aulas de leitura em inglês com corpora para elaboração de materiais didáticos que privilegiem o uso da língua. Esse software foi desenvolvido por Moreira Filho (2007) na sua dissertação de mestrado e está disponível para download em < [corpuslg.org/software/downloads](http://corpuslg.org/software/downloads) >.

<sup>14</sup> Uma das modificações feitas a partir das sugestões do Reading Class Builder, foi a troca de ordem entre as tarefas. O software sugeriu as questões de 2 a 5 após o texto. Optamos por colocá-las após por acreditar que um dos objetivos das perguntas é justamente guiar o leitor para o tipo de informações que devam ser priorizadas. Também foi criada uma parte inteiramente nova, About Academic Abstracts, que privilegia os dados desta pesquisa.

the former Prime Minister of Malaysia, Mahathir Mohamad. This multi-pronged approach will be used to unravel the underlying ideologies for the change and the reluctant compromise that was reached by the Chinese community.

### **Global scientific communication: Open questions and policy suggestions**

Ulrich Ammon; Universität Duisburg-Essen, Germany

This volume is about the role of different languages in science and scientific communication, with the focus on the status and function of entire languages like English, Chinese, Russian, etc. and not on structural details of these languages - following the distinction between status and corpus in language planning. However, the latter i.e. corpus issues (language structure, vocabulary, and rules of pragmatics and discourse), on which the study of languages for special purposes would focus, cannot be entirely ignored as questions of modernization and of terminological and pragmatic adequacy of languages for scientific communication arise in this volume, but have not been dealt with in any detail.

## TEXTUAL FUNCTION ANALYSIS

8) According to Hucking (apud Swales, 2001) abstracts have at least four distinguishable functions:

- A. *They function as stand-alone mini-texts, giving readers a short summary of a study's topic, methodology and main findings;*
- B. *They function as screening devices, helping readers decide whether they wish to read the whole article or not;*
- C. *They function as previews for readers intending to read the whole article, giving them a road-map for their reading;*
- D. *They provide indexing help for professional abstract writers and editors.*

Considering Hucking's ideas above, do you believe the abstract below fulfills these functions?

Read the following abstract and answer the questions.

### **Achievements and challenges in bilingual and multilingual education in the Basque Country**

Jasone Cenoz, *University of the Basque Country*

This paper focuses on the use of Basque as the language of instruction. In the first part of the article the situation of Basque in the Basque Country is briefly described and the different possibilities regarding the language(s) of instruction are explained: model A with Spanish as the language of instruction and Basque as a subject; model B with both Basque and Spanish as languages of instruction and model D with Basque as the language of instruction and Spanish as a subject. Then, the results of research studies comparing these three models regarding achievement in Basque, Spanish and other areas of the curriculum are analysed. Finally the article considers the new challenges the Basque educational system is facing. One of these challenges is the need to go from bilingual education to multilingual education by teaching in a more effective way languages of wider communication. Another recent challenge is multiculturalism as a response to the increasing immigrant population which is a new phenomenon in the Basque educational system. The need for a more holistic approach towards multilingualism both in teaching and research is proposed so as to face these new challenges.

DOI: [10.1075/aila.21.03cen](https://doi.org/10.1075/aila.21.03cen)

In: **Cenoz**, Jasone and Durk **Gorter** (eds.), *Multilingualism and Minority Languages*. 2009 110 pp. (pp. 13-30)

- 9) Underline the extracts which may identify the field and subject of the abstract.
- 10) Does this abstract use citations or references to previous research?
- 11) What is the main tense used in this abstract? Why is this tense used?
- 12) Is the passive voice used? Why?

### ABOUT ACADEMIC ABSTRACTS

In 2011 a research was carried out by Sarmiento et al. about academic abstracts. They analyzed a good amount of data. One category of texts they analyzed was abstracts from TCCs (monographs) and from International Journals on Linguistics.

This is the list of the most frequent words:

Theme →	Linguistics	
<i>Corpus</i> → Words ↓	International Journals	TCCs
1	THE	THE
2	OF	OF
3	AND	IN
4	IN	TO
5	TO	AND
6	A	A
7	<u>LANGUAGE</u>	THAT
8	<u>THAT</u>	<u>THIS</u>
9	AS	IS
10	ON	<u>LANGUAGE</u>
11	FOR	AS
12	THIS	ON
13	IS	FOR
14	ARE	BY
15	WITH	<u>TEACHING</u>
16	<u>ENGLISH</u>	<u>STUDENTS</u>
17	BY	WHICH
18	IT	AN

19	WHICH	FROM
20	AN	WITH

13) What differences and similarities have you found between the two lists?

14) What is special about the underlined words in the table above? How do they differ from the other words which were not underlined?

15) In the following concordance lines, the node word has been erased. What is it? How did you find out?

N	
1	scales of major international _____ tests, as well as in
2	also to refer to those observed _____ learning instances a
3	ly when an intermediate natural _____ is involved in the p
4	n of Ukrainian as the national _____. Based on 10 months
5	ection with language learning, _____ practices, the impac
6	has a compensatory function in _____ attrition, helping t
7	ols have had a central role in _____ revitalisation since
8	rticle examines how inexplicit _____ used by Korean-speak
9	. Building on recent shifts in _____ policy research towa

16) Read the abstract below:

- How similar is it to the two previous one?
- Does it use similar kind of language structures and vocabulary?
- Would you suggest any changes to the abstract below? Which ones? Why?
- Would you like to read it? Why?

This essay contains the report of my teaching practicum which aims to think about Portuguese language teaching and some methodological tenets for its teaching/learning. The study is based on an almost six year experience as a Portuguese teacher at Projeto Educacional Alternativa Cidadã (a popular pré-vestibular course for low class students). This report does not mean to suggest a model, but to prove the importance of a practicum activity in teacher education. The report endeavors to evince the teacher's necessity of pondering on his actions and applying a methodology considering his public and his context. Some essays written by students are presented in order to prove their progress. In the first chapter some theory tenets based on Paulo Freire and Vigotsky are discussed. In the second chapter this six year practice will be reported, as well as the gradual stance changing while the teacher developed a better contact with students' reality. In conclusion, a pondering on how this experience contributed for the development of a new teacher posture and methodology.

## 5. Considerações Finais

Neste trabalho, dividido em duas partes, tivemos por objetivo apresentar a contribuição da Linguística de *Corpus* para a descrição dos gêneros acadêmicos e para a decorrente elaboração de tarefas.

Na primeira parte, a partir da breve análise do *corpus* LILE, detectamos algumas especificidades do gênero *Abstract* e algumas diferenças entre os usos preferidos pelos alunos da graduação e os usos praticados pelos pesquisadores proficientes.

Em relação à segunda parte, tentamos mostrar como uma sequência didática baseada em *corpus* pode ser elaborada. Alertamos que as tarefas devem ser adequadas aos propósitos e interlocutores propostos. No caso em pauta, alunos de graduação do curso de Letras da UFRGS de forma a terem um melhor entendimento do gênero *abstracts*.

Nossos futuros passos nesta pesquisa serão continuarmos com a análise dos outros *subcorpora*, ou seja literatura dos TCCs e Revistas Internacionais e linguística e Literatura de testes e dissertações e Revistas Nacionais. Pretendemos, a partir da análise, elaborar sequências didáticas que realmente ensinem alunos de graduação do curso de Letras a entenderem a função dos *abstracts* e a saberem fazer bom uso do gênero. É também nosso objetivo elaborar tarefas que possam dar conta da habilidade de escrever esse tipo de texto. Após essa etapa, partiremos para a compilação do próximo *corpus*, que será composto de textos completos de artigos, TCCs e teses.

## Referências

- BAKHTIN, M. (2006). *Estética da criação verbal*. São Paulo: Martins Fontes, [1953]
- DAYRELL, C. (2011). *Corpora no Ensino do Inglês Acadêmico: Padrões Léxico-Gramaticais em Abstracts de pós-graduandos brasileiros*. In: VIANA, V.; TAGNIN, S. E. O. *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial Ltda., pp 137-171.
- GRANGER, S; PAQUOT, M. (2007). Spoken features in learner academic writing: identification, explanation and solution. In: *Corpus Linguistics Conference*, 4. Birmingham. Proceeding. Birmingham: University of Birmingham. Disponível em: <[http://www.corpus.bham.ac.uk/corplingproceedings07/paper/204\\_Paper.pdf](http://www.corpus.bham.ac.uk/corplingproceedings07/paper/204_Paper.pdf)>. Acesso em: 15 jun. 2010.
- LEECH, G. N. (1991) *The State of Art in Corpus Linguistics*. London: Longman.
- MCENERY, T. e WILSON, A.(1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- MOREIRA FILHO, J. L. (2007) *Desenvolvimento de um software para preparação de aulas de inglês com corpora*. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem). Faculdade de Comunicação e Filosofia, Pontifícia Universidade Católica de São Paulo, São Paulo.
- O'KEEFFE, A.; MCCARTHY, M.; CARTER, R. (2007) *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.
- SARDINHA, T. B. (2011). Como usar a linguística de *corpus* no ensino de língua estrangeira: por uma linguística de *corpus* educacional brasileira. In: VIANA, V.; TAGNIN, S. E. O. *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial Ltda., pp 301-356.
- SCOTT, M.(2009) *WordSmith Tools. Versão 5.0*. Oxford: Oxford University Press.
- SCHLATTER, M.; GARCEZ, P. (2009) Educação linguística e aprendizagem de uma língua adicional na escola. In: RIO GRANDE DO SUL. Secretaria de Estado da Educação. Departamento Pedagógico. *Referenciais Curriculares do Rio Grande do Sul: Linguagens, Códigos e suas Tecnologias / Secretaria de Estado da Educação*, v.1 Porto Alegre: SE/DP, 2009. p. 127-172.
- STUBBS, M. (2001). *Texts, corpora, and problems of interpretation: A response to Widdowson*. *Applied Linguistics*, 22(2):149-72.

SWALES, J. M. (1990) *Genre analysis: English in academic and research settings*. Nova Iorque: Cambridge University Press.

VIANA, V. (2011) *Linguística de Corpus: Conceitos, Técnicas & Análises*. In: VIANA, V.; TAGNIN, S. E. O. *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial Ltda., pp 19-24.

VIANA, V.; TAGNIN, S. E. O. (2011). *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial Ltda.

### **Periódicos utilizados para a compilação do corpus LILE:**

#### **Literatura- Periódicos Nacionais**

*Literatura e Sociedade*

<http://www.dtlc.fflch.usp.br/revistaliteratura>

*Revista Brasileira de Literatura Comparada*

<http://www.abralic.org.br/revista>

#### **Literatura- Periódicos Internacionais**

*Poetics Today*

<http://poeticstoday.dukejournals.org/current.dtl>

*Language and Literature*

<http://lal.sagepub.com/content/by/year>

#### **Linguística- Periódicos Nacionais**

*DELTA*

[http://www.scielo.br/scielo.php?script=sci\\_issuetoc&pid=0102-445020100001&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_issuetoc&pid=0102-445020100001&lng=en&nrm=iso)

*Trabalhos de linguística aplicada*

[http://www.scielo.br/scielo.php?script=sci\\_issues&pid=0103-1813&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_issues&pid=0103-1813&lng=en&nrm=iso)

#### **Linguística- Periódicos Internacionais**

*AILA Review*

<http://www.aila.info/publications/ailapublications/aila-review/current-reviews.html>

*Applied Linguistics*

<http://appliedjournals.org/content/by/year>