

ANÁLISE DE CORPUS LINGUISTICO EM PESQUISAS DE LEITURA E TRADUÇÃO, FUNDAMENTADA NA LINGUISTICA TEXTUAL DE VAN DIJK E KINTSCH

Marisa Helena Degasperi

RESUMO: O objetivo deste artigo é descrever um modelo de análise de corpus de controle, utilizada em uma investigação científica que resultou em uma tese de doutorado (Degasperi, 2009) sobre processamento de leitura para a produção de resumos em ambiente virtual (AV) e em ambiente não-virtual (ANV). A análise dessa pesquisa considerou as características do texto, fundamentada em teorias da linguística textual de base cognitiva com abordagem discursiva de Teun Adrianus Van Dijk (1978-1996) e Walter Kintsch (1978-2006) que organizam a produção de inferências em três diferentes níveis, de acordo com a organização da estrutura discursiva do texto e o modelo de situação, que adiciona parâmetros semânticos e pragmáticos a esta organização. Segundo esse modelo, se constitui um modelo de situação que deve reger a ordem estrutural do sentido global do texto e requerem diferentes competências do leitor-escritor. Com base na Linguística de Corpus de McEnery & Hardie (2012), nos Estudos da tradução de Orozco e Albir (2002) e Pagano, Magalhães & Alves (2005) faz-se uma adaptação do método para análise de textos em pesquisas com abordagem em leitura para a tradução e em tradução, de análise do TF e do TM, comprovando a viabilidade de sua aplicação para esse tipo de produção textual e também a aposta de aplicabilidade para qualquer outro tipo de texto, com as devidas adequações, em pesquisa que abordam leitura com objetivo específico e em tarefa específica.

PALAVRAS-CHAVE: Leitura; Estudos da tradução; Linguística textual; Modelo situacional.

ABSTRACT: This paper aims to describe a model for an analysis of a control corpus, used in a scientific investigation that resulted in a PhD thesis (Degasperi, 2009) on processing of reading for the production of

summaries in virtual environment (VE) and in non-virtual environment (NVE). The analysis of this investigation considered the text features, based on textual linguistic theories from a cognitive perspective with a discourse approach by Teun Adrianus van Dijk (1978-1996) and Walter Kintsch (1978-2006). These authors organised the production of inferences in three different levels, according to the organisation of text discourse structure and the situation model that adds semantic/pragmatic parameters to this organisation. According to this model, a situation model is constituted, and it has to rule the structural order of the global sense of the text, requiring different competencies of the reader/writer. Based on Corpus Linguistics by McEnery & Hardie (2012), on Translation Studies by Orozco & Albir (2002) and Pagano, Magalhães & Alves (2005), an adaptation is made of the method for the analysis of texts in investigations approaching reading for translation and in translation of the analysis of the source text (ST) and target text (TT). This proved the viability of its application to this type of textual production, and also the relevant assumption for applicability to any other text type, with adequacies, in investigations that approach reading with a specific aim and in a specific task.

KEYWORDS: Reading; Translation Studies; Textual Linguistics; Situational Model.

As pesquisas linguísticas voltadas para os estudos sobre leitura e tradução têm sido consideravelmente ampliadas na última década, principalmente nas abordagens que relacionam as duas tarefas como integradas a um objetivo comum, que é a elaboração, neste caso, de um produto final, o *texto meta* (TM) em uma língua diferente do *texto fonte* (TF) e com o mesmo efeito de sentido (ES) do primeiro no segundo.

Linguística de Corpus e Estudos de Tradução constituem áreas de estudo relativamente recentes e se tornaram mais independentes e interdisciplinares nos últimos anos. (MIGUEL E CHELO, 2013:318).

Segundo Laviosa (2006:8), o corpus constitui-se de dados linguísticos puros que se estabelecem como força motriz da pesquisa.

Nesse sentido, há uma variedade de escolhas potenciais disponível para os investigadores no que se refere aos métodos a serem adotados para sustentar suas propostas teóricas. A análise de corpora linguísticos constituem, obviamente, um dos métodos mais utilizados em pesquisas da categoria citada anteriormente, visto que, tanto o insumo

utilizado como estímulo, quanto o produto final, resultado do processamento de informações, são textos escritos.

O objetivo deste artigo é descrever um modelo de análise de corpus de controle, utilizada em uma investigação científica que resultou em uma tese de doutorado (Degasperi, 2009) sobre processamento de leitura para a produção de resumos em ambiente virtual (AV) e em ambiente não-virtual (ANV) e utilizou como textos fonte dois artigos de uma revista do gênero *popular science*, de circulação nacional. A análise considerou as características do texto, fundamentada em teorias da linguística textual de base cognitiva com abordagem discursiva de Teun Adrianus Van Dijk (1978-1996) e Walter Kintsch (1978-2006) que organizam a produção de inferências em três diferentes níveis, de acordo com a organização da estrutura discursiva do texto. Segundo esse modelo, se constitui um *modelo de situação* que deve reger a ordem estrutural do sentido global do texto e requerem diferentes competências do leitor-escritor. Nesse modelo, se acrescentam outros dois níveis, propostos por Graesser, Millis e Zwaan (1997) *apud* Degasperi (2009):

[...] o nível comunicacional e o nível do gênero do texto. O nível comunicacional faz referência à pragmática da comunicação que o leitor identifica no texto. Neste aspecto, o leitor é forçado a identificar as pistas que levam às referências essenciais do texto, deixadas pelo autor, através de seu estilo. O nível do gênero faz a relação do tipo de discurso com o gênero textual. (DEGASPERI, 2009: 43)

É importante ressaltar de os autores citados fundamentaram sua proposta de níveis inferenciais nas teorias de van Dijk e Kintsch.

A escolha do corpus da pesquisa se deu com base nos princípios de homogeneidade, conforme propõe Bidermann (2001:79): “[...] *corpus* constitui um conjunto homogêneo de amostras da língua de qualquer tipo (orais, escritos, literários, coloquiais, etc.). Tais amostras foram escolhidas como modelo de um estado ou nível de língua predeterminado”; McEnery e Hardie (2012) propõem que a modalidade deva ser em formato eletrônico. Considerou-se, também, a extensão do texto, visto que grande parte, senão a maior parte, das pesquisas sobre leitura e produção de textos, com esse propósito, utilizam apenas palavras, grupos de palavras, frases, fragmentos de textos ou textos muito curtos,

diferente da realidade dos acadêmicos que leem, com maior frequência, artigos ou textos literários completos. Figuras e a superestrutura textual - a forma de organização discursiva e formal do texto-, também formaram parte da análise do texto fonte.

Ao apresentar esse modelo de análise de corpus de uma pesquisa, cujo produto final foi um texto sumarizado, propõe-se, neste trabalho, que o mesmo modelo possa ser utilizado para qualquer tipo de pesquisa em leitura cuja tarefa posterior seja a produção de outro texto, derivado do TF. Essa proposta também descansa no princípio de que o objetivo de qualquer leitura é, *a priori*, a compreensão da estrutura discursiva e a *posteriori*, a utilização das informações do texto em qualquer outra atividade, seja a produção de um novo texto ou a divulgação de dados em outra instância comunicativa. Isto não significa uma marca divisória entre os processos cognitivos. Sendo assim, é possível considerar essa proposta de análise também para a investigação referente à leitura para e na atividade tradutória.

A distinção entre a proposta inicial do modelo para a sua utilização em investigações cuja interface se dá entre leitura e tradução é o objetivo posterior da leitura. Esse objetivo é o responsável pelo planejamento estratégico das atividades (Solé 1996; 1998; Kato, 1999; Kleiman, 1989-2000), dependendo do nível de proficiência do leitor-escritor (Flavell, 1976; Solé 1996; 1998; Kleiman, 1989; Kato, 1999-2000) e, em última análise, resultam na qualidade do texto final, no caso da tradução, o TM.

Contemplando o que propõe a linguística de corpus: o objetivo da análise, nesta perspectiva, de textos escritos, Miguel e Chelo (2013) propuseram analisar uma mostra significativa de 389 corpus linguísticos aplicados à tradução e concluíram que a maior parte dos trabalhos investigativos que utilizam esse método estão dirigidos ao objetivo de investigar os universais da tradução e consideraram que há uma tendência ao conservadorismo em torno da análise puramente linguística, nas pesquisas analisadas. Tendo em vista que a análise de um determinado corpus se subordina aos objetivos da investigação, a projeção que se faz do caráter inovador na investigação em tradução é oposta à manutenção do conservadorismo vigente a partir da abordagem, que também se apresenta diversa nesse âmbito: é o estudo dirigido ao processamento cognitivo das informações. Isto quer dizer que o objetivo é analisar, através do corpus de entrada (TF) e do corpus de saída (TM) os processos previsíveis, padrões (não referente à sequências, pois essas são, ao que parece, impossíveis; mas na frequência de ocorrência), comportamentos derivados

de planejamentos e pressuposições, procedimentos de leitura e de tradução, entre outros aspectos observáveis. Para isso, logicamente, como em qualquer outra metodologia aplicada a pesquisas, o corpus textual será um dos instrumentos que deverão ser utilizados na coleta de dados.

O texto descritivo e os dados apresentados a seguir foram compilados da investigação que produziu esse modelo de seleção e análise de corpus para leitura e produção de resumos, em ambiente virtual (AV) e em ambiente não-virtual (ANV) (Degasperi, 2009). O mesmo modelo está sendo utilizado em um outro trabalho investigativo que está em andamento na Universidade Federal de Pelotas, pela mesma autora, com enfoque nos estudos cognitivos na leitura para e em tradução, com o diferencial de estar atualmente utilizando somente o ambiente virtual, conforme propõem McCnery and Hardie (2012:2):

We could reasonably define corpus linguistics as dealing with some set of machine-readable texts, which is dreamed an appropriate basis on which to study a specific set of research questions. The set of texts of corpus dealt with a usually of a size which defies analysis by hand and eye alone within any reasonable timeframe. Unless we use a computer to read, search and manipulate the data, working with extremely large datasets is not feasible because of the time it would take a human analyst, or team of analysts, to search through the text. It is certainly extremely difficult to search such a large corpus by hand in a way, which guarantees no error.

É importante considerar que não houve interferência da modalidade de ambiente de apresentação do texto na escolha do corpus textual, apenas nos dados resultantes dos demais aspectos considerados naquela pesquisa, que eram relacionados aos comportamentos visuomotor, facial e corporal dos sujeitos, entre outros, que puderam ser contrastados e comparados para efeitos de julgamento de qualidade do produto final. Também foi objetivo daquela investigação estabelecer parâmetros comparativos entre dois ambientes de trabalho.

A proposta é a de utilizar textos completos para as tarefas de leitura e tradução e, por extensão, de análise. E esta proposta se assenta

nas teorias da Linguística textual de abordagem cognitiva-discursiva de van Dijk e Kintsch.

É mister também destacar que as análises do TM poderão ser feitas de acordo com os objetivos da pesquisa e, por assim dizer, sob perspectiva da teoria de preferência do pesquisador. Isto significa que se pode convergir os dados da teoria de base com perspectivas de outras teorias, no que concerne à avaliação do produto final, visto que se podem esgotar a quantidades de informações fornecidas pelos dados.

O Texto fonte

Foram selecionados dois textos da revista *Ciência Hoje* (anexos 1 e 2), com assunto relacionado à área da saúde, que é a área dos cursos dos participantes, para que houvesse familiaridade com as temáticas (Cowen et al, 2001; Richardson, Dale e Spivey, 2007; Drieghe, Desmet, Brysbaert, 2007). Dessa forma, os textos foram considerados de média complexidade, tendo em vista que os assuntos apresentavam termos de alguma área específica, não, necessariamente de algum dos cursos desses sujeitos. Além disso, a fonte de onde se originaram os textos é de circulação nacional com discurso também voltado para o público leigo. Os textos selecionados para os testes, entre 15 textos analisados, foram: o T1 - Alarme arterial - apresentado em ambiente não virtual e o T2 - A cadeira de rodas do futuro? - apresentado em ambiente virtual.

Os textos foram selecionados a partir de uma sistemática de comparação que contemplava os seguintes requisitos de similaridade:

- Temática - o assunto deveria ser relacionado à área da saúde de diferentes áreas dos cursos dos estudantes participantes
- Estrutural - o número de palavras, de linhas e de parágrafos deveria ser muito similar e o formato visual, idem.
- Lexical - os textos deveriam apresentar um conteúdo lexical similar, levando-se em consideração os tipos de palavras que os constituíram.
- Discursiva - as macroproposições deveriam apresentar um esquema muito similar na formação da macroestrutura dos textos, como também o direcionamento do discurso, apresentação de conceitos e definições.

Após análise e seleção dos textos, se obtiveram os resultados, que se apresentam a seguir:

Análise do texto 1	Número caracteres	Número Palavras	Número de frases	Número linhas	Número parágrafos	Número páginas
Título	14	2	1	1	0	0
Subtítulo 1	73	10	1	2	0	0
Subtítulo 2	15	2	1	1	0	0
Texto da Figura	207	33	1	7	1	0
Párrafo 1	650	122	4	11	1	0
Párrafo 2	560	107	4	10	1	0
Párrafo 3	403	80	3	7	1	0
Párrafo 4	395	66	3	7	1	0
Párrafo 5	322	55	2	5	1	0
Párrafo 6	366	70	3	6	1	0
Párrafo 7	344	64	3	6	1	0
Párrafo 8	491	87	3	8	1	0
Total	3840	698	29	71	9	2

Tabela 1 Esquema estrutural e lexical do T1

Análise do texto 2	Número caracteres	Número Palavras	Número de frases	Número linhas	Número parágrafos	Número páginas
Título	24	6	1	1	1	0
Subtítulo 1	70	11	1	1	1	0
Subtítulo 2	15	3	1	1	1	0
Texto da Figura	160	29	1	9	1	0
Párrafo 1	433	86	4	7	1	0
Párrafo 2	413	72	2	7	1	0
Párrafo 3	353	66	3	6	1	0
Párrafo 4	301	59	2	5	1	0
Párrafo 5	350	66	4	6	1	0
Párrafo 6	359	82	2	6	1	0
Párrafo 7	294	52	3	5	1	0
Párrafo 8	392	82	5	6	1	0
Párrafo 9	404	76	6	6	1	0
Total	3568	690	35	66	10	2

Pode-se constatar compatibilidade, ao comparar os dados obtidos de ambos os textos aplicados, respectivamente, em AV e em ANV.

Foram levantados os dados referentes às características lexicais dos textos fonte, no que se refere às funções sintáticas e à semântica. Quanto às funções sintáticas, as palavras foram categorizadas como: funcionais e de conteúdo, sendo as primeiras representadas pelos conectores (preposições, conjunções) e os artigos e referentes dêiticos (pronomes relativos, demonstrativos); palavras de conteúdo, como já se autodenominam, referem-se às palavras que carregam conteúdo semântico, não específico, ou seja, não ligado diretamente à especificidade temática. Os dados obtidos foram os seguintes:

<u>Texto da figura</u>	Nº palavras	
	<u>funcionais</u>	<u>conteúdo</u>
<u>Título</u>	0	2
<u>Subtítulo 1</u>	3	10
<u>Subtítulo 2</u>	2	0
<u>Texto da figura</u>	9	24
<u>Parágrafo 1</u>	53	69
<u>Parágrafo 2</u>	43	64
<u>Parágrafo 3</u>	29	51
<u>Parágrafo 4</u>	19	47
<u>Parágrafo 5</u>	25	30
<u>Parágrafo 6</u>	22	48
<u>Parágrafo 7</u>	23	41
<u>Parágrafo 8</u>	34	53
Total	262	439

Tabela 3 Estrutura sintática do texto 1

<u>Texto 2</u>	Nº palavras	
	<u>funcionais</u>	<u>conteúdo</u>
<u>Título</u>	3	3
<u>Subtítulo 1</u>	3	8
<u>Subtítulo 2</u>	1	2
<u>Texto da figura</u>	10	19
<u>Parágrafo 1</u>	33	53
<u>Parágrafo 2</u>	24	48
<u>Parágrafo 3</u>	23	43
<u>Parágrafo 4</u>	23	36
<u>Parágrafo 5</u>	22	44
<u>Parágrafo 6</u>	29	53
<u>Parágrafo 7</u>	20	32
<u>Parágrafo 8</u>	32	50
<u>Parágrafo 9</u>	23	53
Total	246	444

Tabela 4 Estrutura sintática do texto 2

Também neste quesito, podem-se perceber as similaridades entre ambos os textos fonte.

As palavras de conteúdo específico também foram analisadas. Por conteúdo específico entende-se o conteúdo semântico diretamente ligado à temática do texto. Neste caso, considerou-se a extensão das palavras. Por “palavras curtas” entendam-se palavras com até 4 sílabas, por “médias”, de 5 a 7 sílabas e por “palavras longas” as de 8 ou mais sílabas. Esta classificação foi feita para contribuir, se necessário, com os dados das filmagens dos movimentos oculares. Os critérios de classificação das palavras, nesses aspectos foram os seguintes:

As palavras compostas, separadas por hífen, quando não apresentavam palavras funcionais (i.e. micro-hemorragias, médico-científica); palavras funcionais (i.e. tê-los) eram consideradas uma só e, se

a palavra funcional aparecesse entre duas palavras de conteúdo, era desconsiderada (i.e. cadeira de rodas).

O símbolo de porcentagem foi agregado ao número formando uma só palavra.

Nomes próprios tiveram todas as palavras agregadas, contando como uma só palavra, considerando o significado e a representação serem um só.

Todos os verbos que se relacionavam especificamente ao tema do texto foram considerados palavras específicas. Os resultados se apresentam a seguir:

Texto 1	Palavras	Palavras	Total de
	específicas	específicas	palavras
	longas	médias/curtas	específicas (Parágrafo)
Título	1	0	1
Subtítulo 1	2	5	7
Subtítulo 2	0	0	0
Texto da figura	5	8	13
Parágrafo 1	18	11	29
Parágrafo 2	16	13	29
Parágrafo 3	7	14	21
Parágrafo 4	12	5	17
Parágrafo 5	5	6	11
Parágrafo 6	10	11	21
Parágrafo 7	19	11	30
Parágrafo 8	19	6	25
Total palavras			
	114	90	204
específicas (texto)			

Tabela 5 Estrutura semântica do T1

Texto 2	Palavras	Palavras	Total de
	específicas	específicas	palavras
	longas	médias/curtas	específicas (Parágrafo)
Título	2	1	3
Subtítulo 1	4	4	8
Subtítulo 2	2	1	3
Texto da figura	5	7	12
Parágrafo 1	12	12	24
Parágrafo 2	17	10	27
Parágrafo 3	14	10	24
Parágrafo 4	10	13	23
Parágrafo 5	12	11	23
Parágrafo 6	9	17	26
Parágrafo 7	9	12	21
Parágrafo 8	13	14	27
Parágrafo 9	14	8	22
Total palavras específicas (texto)	123	120	243

Tabela 6 Estrutura semântica do T2

Nesta condição, há uma diferença de 39 entre palavras específicas a mais para o T2 em relação ao T1, pelo que se poderia conjecturar que o T2 pudesse ser considerado pelos estudantes como mais difícil que o T1, ou de linguagem mais técnica. Porém, a denominação específica não se refere somente a termos técnicos, senão, como já foi citado anteriormente, palavras que se relacionam diretamente com a pesquisa (i.e. técnica, eficiência, sangue, morte, sistema, etc.), que podem ser consideradas, também, palavras de significado facilmente identificável. Em resumo, o fato de serem palavras específicas não significa que

constituam dificuldades para o leitor, senão que possuem um potencial de dificuldade semântica maior que as não específicas.

A semelhança do conteúdo visuoestrutural dos textos também pode ser verificada a seguir:

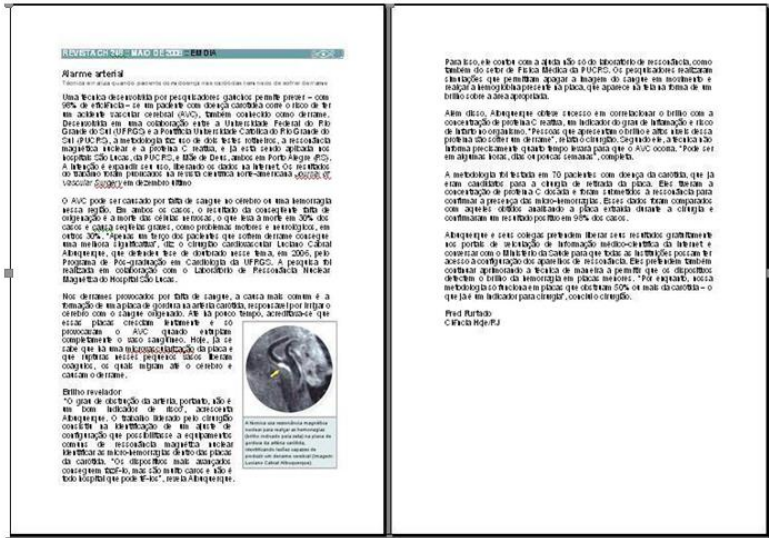


Imagem 1. Conteúdo visuo-estrutural do T 1

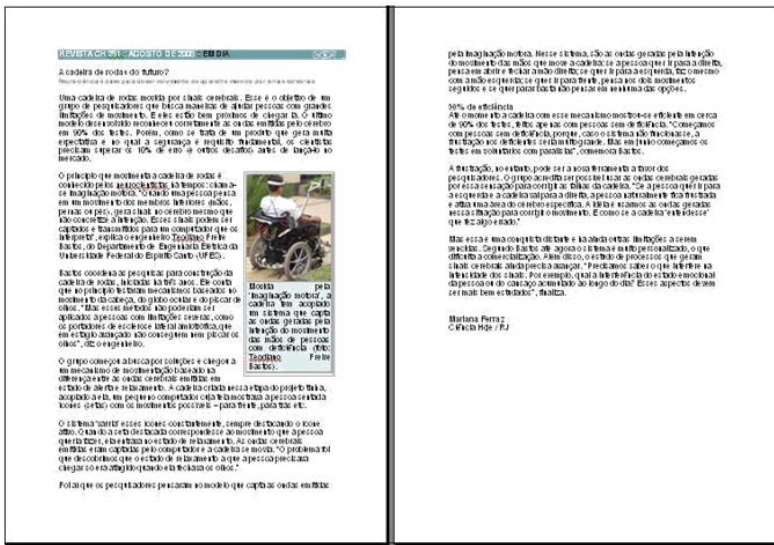


Imagem 2. Conteúdo visuo-estrutural do T 2

Como se pode perceber, os conteúdos dos textos, no que se refere aos aspectos estruturais e visuais não apresentam diferenças que se possam considerar relevantes.

Quanto ao aspecto discursivo, analisaram-se os tópicos temáticos por parágrafo (macroproposições) e os tópicos frasais (microestruturas), para que seja possível a análise dos rascunhos e dos resumos, quando da avaliação da produção, que se poderão verificar no anexo

Os tópicos frasais considerados de maior relevância informacional na construção das macroproposições e, consequentemente, para o contexto situacional originado da macroestrutura textual foram selecionados e aparecem em destaque. Esses tópicos foram os que nortearam, também, a avaliação dos textos resumo dos participantes.

ANÁLISE DO CONTEÚDO DOS TEXTOS

Texto 1

Divisão do assunto texto por parágrafos (macroproposições):

1º parágrafo: Introdução/ resumo do conteúdo - técnica, objetivo, referência do lugar; colaboradores, descrição e universo da aplicação, objetivos e resultados.

- 2º parágrafo: descrição AVC, estatísticas, autor, colaboradores.
- 3º parágrafo: causas dos derrames e descobertas recentes
- 4º parágrafo: metodologia utilizada e limitações
- 5º parágrafo: departamentos envolvidos na pesquisa e técnicas utilizadas
- 6º parágrafo: descobertas limitações
- 7º parágrafo: população da pesquisa, operacionalização dos dados e resultados
- 8º parágrafo: perspectivas - projeções futuras e limitações da metodologia.

Divisão do assunto texto por frases (conteúdo semântico - tópicos frasais):

1. **título**
2. subtítulo 1
3. **a técnica, estatísticas e objetivo, explicação;**
4. onde se desenvolve e instituições, tipos de testes e aplicação;
5. objetivo dos pesquisadores
6. publicação dos resultados - periódico e data;
7. causas do AVC;
8. consequências do AVC;
9. estatísticas; apresentação acadêmica do cirurgião que liderou as pesquisas;
10. instituições colaboradoras;
11. **causa comum de derrames;**
12. conclusões anteriores;
13. **conclusões atuais;**
14. obstrução não bom indicador
15. **técnica empregada;**
16. impossibilidade de utilização de dispositivos mais adiantados (caros);
17. departamentos e instituições colaboradoras;
18. **técnica - simulações e detecção das micro hemorragias;**
19. operacionalização dos dados - correlação das variáveis
20. **resultado da aplicação do teste - prognóstico;**
21. imprecisão da técnica, limitação;
22. previsão - sem precisão

23. população da pesquisa;
24. aplicação da metodologia na população da pesquisa;
25. comparação dos dados;
- 26. pretensão de divulgação dos resultados em setores de divulgação;**
- 27. pretensão de aprimoramento da técnica;**
- 28. limitação da metodologia, da técnica e dos testes empregados.**

Texto 2

Divisão do assunto texto por parágrafos (macroproposições):

- 1º parágrafo: Introdução - objeto do texto, objetivo dos pesquisadores, o que desenvolveram, resultados estatísticos e limitações da pesquisa;
- 2º parágrafo: conceito de imaginação motora apresentação do engº responsável;
- 3º parágrafo: histórico da pesquisa - antecedentes e situação problema;
- 4º parágrafo: busca de soluções e descrição material da cadeira;
- 5º parágrafo: funcionamento da cadeira
- 6º parágrafo: aperfeiçoamento do sistema e novas operações da cadeira;
- 7º parágrafo: resultado estatístico e cautela nos testes;
- 8º parágrafo: a frustração como ferramenta favorável à pesquisa na correção de falhas.
- 9º parágrafo: Limitações e projeções futuras de avanço da pesquisa.

Divisão do assunto texto por frases (conteúdo semântico – tópicos frasais):

- 1. título;**
2. subtítulo 1;
- 3. o objeto do texto: a cadeira;**
- 4. objetivo do grupo de pesquisadores;**
5. projeção otimista;
- 6. modelo mais recente e estatísticas positivas;**
- 7. intenção de aperfeiçoamento da cadeira;**
8. princípio que movimenta a cadeira: imaginação motora;
9. definição de imaginação motora;
10. possibilidade de utilização da imaginação motora e apresentação do engenheiro responsável pela pesquisa;
11. coordenação e tempo da pesquisa;
12. experiências anteriores, detalhes;
13. problemas apresentados;

14. busca de soluções do grupo;
15. descrição do funcionamento - física; movimentos;
16. descrição do funcionamento - sistema;
17. descrição do funcionamento - estado de relaxamento;
18. descrição do funcionamento - ondas cerebrais;
19. descoberta do problema;
20. aperfeiçoamento do modelo anterior;
21. **novas técnicas e funcionamento;**
22. eficiência da nova tecnologia aplicada à cadeira de rodas;
23. **eleição de sujeitos normais como forma de evitar frustrações;**
24. testes em voluntários deficientes;
25. Frustração como ferramenta a favor dos pesquisadores
26. **uso da frustração para corrigir falhas;**
27. explicação da frustração;
28. **ideia de usar ondas geradas a partir da frustração**
29. explicação desse uso - como a cadeira “reagiria”;
30. **projeções futuras e limitações;**
31. **dificuldades de comercialização;**
32. **necessidade de avanços nas pesquisas;**
33. questionamento sobre interferência nos sinais;
34. questionamento sobre estado emocional e cansaço;
35. necessidade de mais estudos sobre o tema.

O esquema discursivo dos textos também foi identificado e comparado e apresentou-se, em ambos, com o seguinte formato:

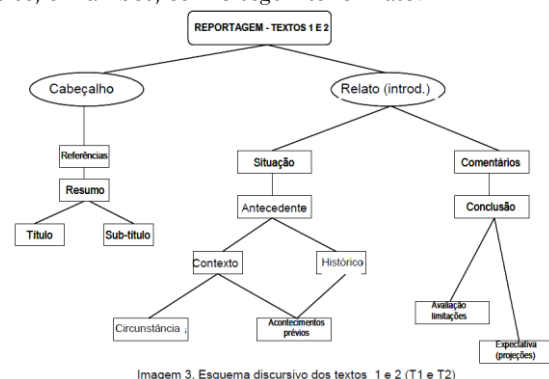


Imagem 3. Esquema discursivo dos textos 1 e 2 (T1 e T2)

Não houve, como se pode constatar, diferenças na estrutura discursiva dos textos fonte, o que comprova, novamente, a compatibilidade de ambos.

O texto fonte apresentado em ANV tinha o formato de fotocópia em cores, da mesma forma como o texto fonte em AV. Neste ambiente, o texto foi apresentado em formato de texto corrido (movimento do mouse para mudança de páginas).

Avaliando o produto final – texto resumido

Os critérios para avaliação de cada um dos quesitos foram fundamentados nos requisitos apontados por Charolles (1991) – brevidade, fidelidade, coesão e coerência; em van Dijk e Kintsch (1983) – conjunção dos níveis micro e macroestrutural que estabelecem a coerência e coesão textual e a informatividade e o manejo das informações, através dos conectores e elementos redutores que constituem a coesão textual (Vigner, 1991).

As microestruturas dos textos foram estabelecidas pelos tópicos frasais do texto fonte. Os critérios fixados para a avaliação dos textos resumos foram os seguintes: fidelidade, informatividade, produtividade e redução (brevidade).

Observe-se que, ainda que em ambos os ambientes, o limite estipulado de linhas foi de 15, considerou-se o número de palavras como referência para avaliar o quesito redução, que representa a brevidade do texto comparando-se com o texto fonte.

A coerência da produção foi indicada conforme critérios de fidelidade e organização das informações do texto fonte (Charolles et al., 1986) e foi avaliada através da verificação da compatibilidade das asserções com aquele texto e o grau de informatividade, conforme já foi explicado anteriormente. A coesão, neste caso, está representada pela condensação das ideias relevantes do texto fonte, através do uso de conectores e de sua hierarquia e redução do texto.

Percentual	Grau
0- 49%	1
50-79%	2
80-100	3

Níveis do grau de informatividade do texto resumo

A avaliação do texto resumo como produto final dos processamentos de leitura, sumarização e escritura teve como base as teorias citadas anteriormente, referentes às qualidades do resumo. Para essa avaliação, utilizaram-se, as seguintes categorias: Fidelidade; Informatividade; Produtividade; Brevidade ou redução do número de palavras.

Para avaliar o nível da qualidade final do texto resumo utilizaram-se três níveis que concentram os resultados de cada uma das categorias definidas anteriormente, com base nos graus utilizados por Anglat (2008).

Níveis de qualidade do resumo	
Independente - 3	Entre 2,9 e 3,0
Instrucional - 2	Entre 2 e 2,9
Dependente - 1	Entre 1 e 1,9

Critérios de avaliação da qualidade do resumo em níveis

Anglat (2008) estabeleceu estes níveis para avaliar fluência leitora, precisão do conteúdo e compreensão leitora, com base em pressupostos de Condemarin e Milicic (1988).

O nível independente supõe domínio metacognitivo também mencionado por Kintsch e van Dijk (1983); o instrucional domínio parcial do processamento; e o de frustração, substituído aqui por “dependente”, sugere uma superficialidade da leitura e incapacidade de integrar os níveis de compreensão apontados por Kintsch e Dijk (1983). Estes três níveis condensam os resultados da análise dos elementos constitutivos de cada uma das categorias citadas anteriormente:

1. Fidelidade: qualidades de V (verdadeira – informação consistente) ou F (falsa – informação inconsistente ou inexistente no texto fonte) para cada informação do texto resumo em relação às informações do texto fonte (microproposições). Ao final, faz-se um cálculo da porcentagem de informações verdadeiras e falsas, com os seguintes parâmetros:

Percentual	Grau de fidelidade
0- 49%	1
50-79%	2
80-100%	3

Níveis do grau de fidelidade do texto resumo

Então, quanto mais informações verdadeiras, mais fiel é o texto resumo.

2. Informatividade: informa a quantidade de informações oferecidas pelo texto, em relação às macroproposições do texto fonte (Charolles 1991: p.48-49). Dessas informações, serão consideradas as de maior relevância, representadas pelos tópicos frasais destacados na tabela anterior.

Quanto maior o número de microproposições relevantes, maior a informatividade. Os mesmos parâmetros utilizados para a fidelidade foram utilizados para a informatividade, da porcentagem do nº de informações que constitui o texto resumo, compatíveis com as informações relevantes do texto fonte.

Percentual	Grau
0- 49%	1
50-79%	2
80-100	3

Níveis do grau de informatividade do texto resumo

Para esta qualidade do texto, quanto maior o número de informações relevantes do texto fonte utilizadas no texto resumo, maior a informatividade.

3. Produtividade: evidenciada pela construção de sentenças através de agrupamento de tópicos frasais relevantes ou de interpretação, representada pela produção de sentenças completamente diferentes do texto fonte, mas coerente com a macroestrutura original. (van Dijk, 1978-1980; Charolles, 1983; Smith, 1983; van Dijk & Kintsch, 1983).

Foi medida seguindo os seguintes critérios de produção:

- a. *copy-delet* e paráfrase (valor atribuído: 1 - menos produtivo)
- b. agrupamento de informações (valor atribuído: 2 - produtivo)
- c. interpretação (valor atribuído: 3 - mais produtivo)

Não se atribuiu valor 0 (zero) para nenhuma das formas de produção por julgar que, mesmo utilizando estratégia de menor esforço cognitivo, o sujeito precisou fazer seleção e organização das informações, por menor que tenha sido seu trabalho.

A valoração numérica final seguiu o mesmo modelo dos anteriores, para haver coerência nos dados finais:

Percentual	Grau
0- 49%	1
50-79%	2
80-100	3

Níveis do grau de produtividade do texto resumo

4. Brevidade ou redução do número de palavras: esta categoria não se evidencia como as demais qualidades avaliadas de maneira isolada; somente no conjunto esta característica pode evidenciar a sintetização do texto fonte. (Charolles, 1986; Vigner, 1991). Para esta qualidade do resumo, em relação ao texto fonte, utilizaram-se as seguintes medidas: Então, quanto maior a redução, maior a brevidade do texto, recordando que esta qualidade do texto é dependente das demais qualidades, isto é, um texto mais reduzido não é, necessariamente, o melhor.

Percentual de redução	Grau
40 - 50	3
30- 39	2
10 - 29	1

Níveis do grau de brevidade ou redução do texto resumo

Entre os processos estudados, a produção do texto resumo, por assim dizer, é uma atividade que requer, como em qualquer atividade de escritura: planejamento, automonitoramento, avaliação e revisão final do texto. Entretanto, como já se disse anteriormente e é de conhecimento geral, é também uma atividade que requer um esforço cognitivo muito grande, visto que há vários processos envolvidos em seu desenvolvimento. O resumo desponta, então, como produto final dos processamentos de leitura e sumarização do texto fonte e também como instrumento de compreensão leitora.

A qualidade dos resumos produzidos pelos participantes, em ANV e em AV, foi avaliada com base na utilização das macrorregras que guiam o processamento da sumarização, sugeridas por van Dijk e Kintsch (1983) e nas qualidades do resumo, como texto sumarizado, listadas por van Dijk (1978), van Dijk e Kintsch (1983), Charolles (1991) e Vigner (1991), para as quais convergem os esforços cognitivos da produção desse tipo de texto.

Fazendo uma breve retrospectiva, para avaliar o nível da qualidade final do resumo, para o conjunto das categorias, utilizaram-se três níveis, que concentram os resultados de cada uma das categorias definidas anteriormente, com base nos níveis definidos por Anglat (2008), substituindo-se o termo frustração, utilizado pela autora, pelo termo dependente:

Níveis de qualidade do resumo	
Independente - 3	Entre 2,9 e 3,0
Instrucional - 2	Entre 2 e 2,9
Dependente - 1	Entre 1 e 1,9

Níveis de qualidade de resumos

Adequação da proposta a pesquisas com interface entre leitura e tradução

O modelo proposto tem enfoque na produção de resumos acadêmicos e a proposta deste artigo tem a ótica voltada para a tradução. Dessa forma, pode-se afirmar que é possível adequar a análise de corpus linguístico para o texto-fonte utilizando os mesmos parâmetros utilizados, do ponto de vista estrutural, lexical, sintático e semântico, bem como superestrutural em diferentes línguas. Por outro lado, a avaliação do produto final não poderia ser realizada da mesma maneira, visto que a tipologia textual final é diversa, bem como os processos durante as atividades, são bastante diferentes. Isto é relevante porque a funcionalidade de ambas as produções pleiteiam diferentes objetivos. Sendo assim, é certo que estamos tratando de diferentes processamentos de informação com diferentes demandas cognitivas e, portanto, o tratamento do seu resultado final requer outros quesitos.

O que pode ser aproveitado, então, dos quesitos apresentados para o resumo na avaliação da tradução? O nível de adequação e de similaridade do efeito cognitivo do TF no TV é, obviamente, o princípio que fundamenta os dois tipos de produção textual.

Sugestão de critérios de avaliação para traduções

A abordagem da análise poderá ser *qualitativa e quantitativa*, por entender que uma pesquisa com apenas uma dessas abordagens seria muito difícil ou mesmo impossível no momento da análise de dados:

“Concordances and frequency data exemplify respectively the two forms of analysis, namely qualitative and quantitative, that are equally important to corpus linguistics.” (MCENERY & HARDIE, 2012:2)

É importante ressaltar que o modelo apresenta uma proposta de análise e avaliação que não pressupõe ser a única nem ser inquestionável, mas se legitima pela cientificidade da eleição de aspectos no produto final, o TM, que é analisável, palpável e torna válidos os graus e níveis designados para esta avaliação.

Quesitos a ser considerados: Adequação lexical, sintática, semântica e pragmática. Esses quesitos poderão revelar o grau de eficiência tradutória, que identificarão a competência *do trabalho* do

tradutor (e não o tradutor), que poderá ser determinado, alcançando-se uma média ponderada de todos os níveis. A escolha desses quesitos se baseia na *competência tradutória* nestes aspectos ao julgá-los essenciais para o alcance do *efeito de sentidos* proposto pela tarefa tradutória.

O conceito utilizado para *competência tradutória* para os níveis se fundamentam nas teorias de (Pagano, Magalhães & Alves, 2005; Orozco e Albir, 2002), que envolvem habilidades tradutoras e competências (e subcompetências) que conduzem à expertise e ao sucesso da atividade tradutória. Por outro lado, o conceito de eficiência, foi adotada com base na *competência textual* que envolve a visão psicológica e linguística do texto de van Dijk (1978-1981) e a reconstrução das informações, com caráter de fidelidade, nos níveis micro, macro e supertextuais.

Níveis de competência tradutória

Muito competente	95 a 100%
Competente	85 a 95%
Pouco competente	70 a 85%
Insuficientemente competente	- 70%

Qualificações dos quesitos:

Palavras de conteúdo - Número de palavras e % de palavras inadequadas (que incorra a erro semântico/ pragmático). Incluir termos, quando o texto for técnico. A fórmula a ser utilizada, seria $N^{\circ} PA - N^{\circ} PI = N^{\circ} PE \Leftrightarrow \%PA - \%PI = \% E$

Muito Eficiente	95 a 100%
Eficiente	85 a 95%
Pouco eficiente	75 a 85%
Ineficiente	- 75%

Palavras funcionais - Número de palavras adequadas e % de palavras inadequadas. Considerando que os conectores, atributivos e designações de seres são essenciais para a coerência discursiva, teremos. $N^{\circ} PA - N^{\circ} PI = N^{\circ} PE \Leftrightarrow \%PA - \%PI = \% E$

Muito Eficiente	95 a 100%
-----------------	-----------

Eficiente	85 a 95%
Pouco eficiente	75 a 85%
Ineficiente	- 75%

Organização sintática - Número de tópicos frasais e % de inadequações (que incorram em distorção, inconsistência ou incompletude de sentido)

Uma forma de estabelecer estatisticamente as microestruturas das proposições é indicar o grau de informatividade da proposição como: informação completa, informação incompleta e informação distorcida ou inconsistente, atribuindo para esses quesitos os graus C (completa), I (incompleta) e D (distorcida ou inconsistente), respectivamente e verificar, ao final, a seguinte fórmula: $N^{\circ} C - N^{\circ} D = \% C - \% D \rightarrow \% C + \% I / 2 = \% E$ (média ponderada de eficiência)

Nº de informações completas - % de adequações em relação a todas as informações do texto;

Quanto maior a % de informações adequadas, mais eficiente será a tradução.

Qualidade	%
Muito Eficiente	90 a 100%
Eficiente	80 a 90%
Pouco eficiente	-80%

Conjectura-se, no caso do quesito organização sintática, que 20% de inadequação no número de proposições, causa de distorções, inconsistência ou incompletude dos efeitos de sentido do TF no TM e podem comprometer o conteúdo em nível de informação global dos parágrafos e tornar o trabalho tradutório final, ou TM, ineficiente.

Adequação semântica e pragmática - macroestrutura sem sequência lógica ou transposição de informações não fiéis ao sentido do TF. Número de proposições e % de desconexão. Para esse quesito pode-se atribuir os rótulos: A (adequado) e I (inadequado).

Neste caso, quanto maior for a quantidade de informações adequadas, maior será a eficiência do TM. O resultado poderá ser obtido com a seguinte fórmula: $\%A - \%I = \% E$

Muito Eficiente	95 a 100%
Eficiente	80 a 95%
Pouco eficiente	-80%

Supõe-se, no caso do quesito semântica e pragmática, que 20% de inadequação no número de proposições seria bastante prejudicial para os efeitos de sentido do TF no TM, considerando-se, assim, uma ineficiência que torna desfavorável a indicação do texto final como tradução.

Em todos esses casos, concebe-se a porcentagem de adequações como positiva, subtraídas as porcentagens de inadequações. Retomando ao resultado qualificativo do trabalho final da tarefa tradutória, já indicado anteriormente e recordando que não se estabelecem juízos sobre a competência geral do tradutor, mas dos trabalhos analisados, tem-se:

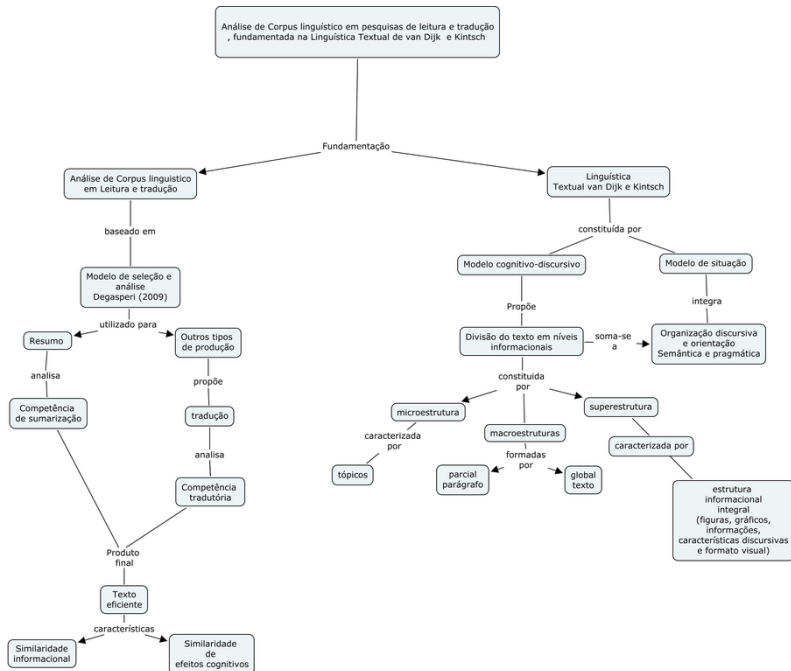
Níveis de competência tradutória do texto (1,2,3 ou A,B,C)

Somatória dos níveis de eficiência (NE) alcançados nos quesitos anteriores em $\%NE1+\%NE2+\%NE3+\%NE4/4 = NC$ (Nível de competência tradutória)

Qualidade	Nível	Alcance de eficiência
Muito competente	4	95 a 100%
Competente	3	85 a 95%
Pouco competente	2	70 a 85%
Insuficientemente competente	1	- 70%

Como também já se estabeleceu anteriormente, este é um modelo de análise proposto que admite diferentes rotulações e graduações e níveis, dependendo do objetivo do pesquisador e se dispõe unicamente à contribuição como método de análise de trabalhos de tradução.

Melhor representação mental do método de análise proposto neste artigo pode ser visualizado no esquema a seguir:



A proposta para a pesquisa em leitura para e em tradução do projeto atual de DEGASPERI (UFPel, 2014) é a de análise de um corpus monolíngue para os textos fonte (TF), cujas traduções serão vertidas ao português Tradução L1 e um corpus monolíngue para a atividade de Tradução L2. Os textos meta (TM) constituirão, então, o corpus bilíngue para análises relacionadas ao processamento tradutório, tendo em conta que são processamentos de diferentes línguas, por diferentes sujeitos. Toda distinção entre os elementos que participam das atividades deve ser considerada dentro de sua especificidade: sujeito-leitor – que carrega suas experiências de mundo e suas competências; o TF – que está escrito, originalmente em determinada língua e tem uma funcionalidade, carrega o estilo do autor, que também é um sujeito; o TM – que é o resultado de diferentes processos cognitivos do sujeito leitor-tradutor e os demais elementos de análise. A relevância de todos os aspectos que envolvem o processamento cognitivo e de informações em diferentes línguas denotam

a complexidade das tarefas, leitora e tradutória e provocam reflexão acerca da limitação da simplificação dos processos e sua classificação em inventários de procedimentos e de resultados linguísticos concretos.

A proposta apresentada na pesquisa anterior (Degasperri, 2009), evidencia que o modelo pode ser utilizado de maneira prática, para a análise de corpus de TF em tradução, com as devidas adequações e pode constituir-se, assim, uma nova perspectiva na pesquisa em leitura para e em tradução, evitando, assim, a rigidez do conservadorismo.

BIBLIOGRAFIA CONSULTADA

ALVES, F. **Ritmo cognitivo, meta-reflexão e experiência: parâmetros de análise processual no desempenho de tradutores novatos e experientes.** In: PAGANO, A.; 2266 Curitiba 2011 Anais do VII Congresso Internacional da ABRALIN.

ANGLAT, Hilda DiFabio. El test cloze en la evaluación de la comprensión del texto informativo de nivel universitario. **Rev de Ling Teórica e Aplic**, I sem. Concepción (Chile): 2008. p 121-137

BAKER, M. Corpus in Translation Studies: an overview and some suggestions for future research. In: **Target 7:2**. Amsterdam: John Benjamins, 1995.

BIDERMANN, M.T.C. **Teoria Linguística**. 2. ed. São Paulo: Martins Fontes, 2001.

CHAROLLES, M. Lê résumé de texte scolaire: fonctions et principes de élaboration. **Pratiques**, Meta, n. 72, dé. 1991.

CHAROLLES, Michael; COLTIER, Danielle. Le controle de la compréhension dans une activité rédactionnelle: éléments pour l'analyse de la reformulation paraphrastique. **Pratiques** n° 49, mars 1986, pp. 51-66

CONDEMARÍN, M., & MILICIC, N. Test de Cloze: **Procedimiento para el desarrollo y la evaluación de la comprensión lectora**. Editorial Andrés Bello. Santiago-Chile:1988.

COWEN, Laura; Ball, LINDEN, J.; DELIN, Judy. An eye movement analysis of web-page usability. Department of Psychology, Lancaster University, Lancaster, UK: 2001.

DEGASPERI, M. H. Processamento de leitura e de produção de resumos em ambiente virtual e em ambiente não-virtual. Tese de doutorado. PUCRS. Porto Alegre: 2009.

DIJK, T. A. Van. **La ciencia del texto**. Traducción de Sibila Hunzinger. Paidós. México: 1978.

_____. Macrostructures. An interdisciplinary study of global structures in discourse, interaction, and cognition. Hillsdale, NJ: Lawrence Erlbaum: 1980.

_____. **Cognição, Discurso e Interação**. São Paulo: Contexto, 1996.

_____. **Estructuras y funciones del discurso: una introducción interdisciplinaria a la lingüística del texto y a los estudios del discurso**. Trad. de Myra Gann e Marti Mur. Siglo Veintiuno Editores. México: 1996.

DRIEGE, Denis; DESMET, Timothy; BRYSSBAERT, Marc. How important are linguistic factors in word skipping during reading? **British journal of psychology**, vol. 98 (1): 2007. p. 157-171

FERRAZ, Mariana. A cadeira de rodas do futuro? Neurociência é base para desenvolvimento de aparelho movido por sinais cerebrais. **Revista Ciência Hoje**, 251. SBPC UFRJ. Rio de Janeiro. Agosto de 2008.

FLAVELL, J.H. Metacognitive aspects of problem solving. In: RESNICK, L. (Ed.): **The nature of intelligence**. Hillsdale, LEA: 1976.

_____. Speculation about the nature and development of metacognition. In WEINERT, F.E. y KLUWE. R.H. (Eds.): **Metacognition, motivation and understanding**. Hillsdale: Erlbaum: 1987. p. 21-30.

FURTADO, Fred. Alarme arterial: Técnica sinaliza quando paciente com doença nas carótidas tem risco de sofrer derrame. **Revista Ciência Hoje**, 251. SBPC UFRJ. Rio de Janeiro. Agosto de 2008.

GRAESSER, Arthur R; MILLIS, KEITH K; ZWAAN Rolf A. Discourse and comprehension. **Annual Review of Psychology**, Vol. 48, 1997.

KINSTCH, Walter. The role of knowledge in discourse comprehension: a construction-integration model. **Psychological Review**. Colayright by the American Psychological Association, Inc. 1988, Vol. 95, No. 2, 163-182KATO, Mary. **O aprendizado da leitura**. 5ª ed. São Paulo: Martins Fontes, 1999.

KINTSCH, W, DIJK, T. A, V. Strategies in discourse comprehension. New York: Academic Press: 1983.

- KINSTCH, Walter; RAWSON, Katherine A. Comprehension. In: **The Science of Reading: A Handbook** por Margaret J. SNOWLING, Charles HULME. Blackwell Publishing, 2005, p.209-226.
- KLEIMAN, A. **O texto e o leitor: aspectos cognitivos da leitura**. Pontes, Campinas, 1989.
- _____. **Oficina de leitura: teoria e prática**. 7ª ed. São Paulo: Pontes, 2000.
- LAVIOSA, Sara. The corpus based-approach: a new paradigm in translation studies. Special issue of **Meta** 43 [4]:1998, p. 557-570.
- MAGALHÃES, C.; ALVES, F. (Org.). **Competência em tradução: Cognição e Discurso**. Belo Horizonte: UFMG, 2005b. p. 109-169.
- MCENERY Tony, HARDIE, Andrew. **Corpus Linguistics: Method, Theory and Practice. Cambridge textbooks and linguistics**. Cambridge University Press. United Kingdom: 2012.
- MIGUEL A. Candel-Mora, CHELO, Vargas-Sierra / *Procedia - Social and Behavioral Sciences* 95 (2013) 317 – 324
- OROZCO, M. & HURTADO ALBIR, A. Measuring translation competence acquisition. In **META**, n. 47, v. 3, p. 375-402, 2002.
- PACTE. La competencia traductora y su adquisición, **Quaderns**, 6, 2001, p.39-45.
- RICHARDSON, Daniel C.; DALE, Rick; SPIVEY, Michael J. Eye movements in Language and Cognition: a brief introduction. In **Methods in Cognitive Linguistics** Edited by Monica GONZÁLEZ-MÁRQUEZ, Irene MITTELBERG; Seana COULSON; Michael J. SPIVEY .Cornell University / UCSD / Cornell University: 2007. 323-344
- SMITH, Frank. Reading like a writer. **Language Arts, Urbana**, National Council of Teachers of English, v. 60, n. 5, p. 558-67, may 1983.
- VIGNER, Gérard. Réduction de l'information et généralisation: aspects cognitifs et linguistiques de l'activité de résumé, pp. 33-54. **Pratiques**. N.72, p. 33-54: 1991.

Tradução inversa do abstract deste artigo de Roberta Rego Rodrigues.