

DOI: <http://dx.doi.org/10.1590/1982-4017-25-03>

Recebido em: 12/07/2024 | Aprovado em: 09/12/2024

Artigo Original

Editor de Seção: Fábio José Rauhen

REVISÃO SISTEMÁTICA DE DEFINIÇÕES E CARACTERIZAÇÕES DE TEXTO: UMA PROPOSTA DE DEFINIÇÃO OPERACIONAL

Systematic Review of Definitions and Characterizations of Text: A Proposal for an Operational Definition	Revisión sistemática de definiciones y caracterizaciones de texto: una propuesta de definición operacional
--	--

Davi Alves Oliveira*

Universidade do Estado da Bahia,
Departamento de Ciências Humanas – Campus IV, Jacobina, BA, Brasil

Lêda Maria Braga Tomitch**

Universidade Federal de Santa Catarina,
Departamento de Língua e Literatura Estrangeiras, Florianópolis, SC, Brasil

Hernane Borges de Barros Pereira***

Universidade do Estado da Bahia,
Departamento de Educação – Campus I, Salvador, BA, Brasil
Centro Universitário SENAI CIMATEC, Salvador, BA, Brasil

Resumo: Este estudo investiga respostas para a complexa pergunta “o que é texto?” e propõe uma definição operacional de texto que sintetize as características gerais encontradas e especifique algumas características da materialidade textual. Através de uma revisão sistemática de artigos científicos, diversas definições e caracterizações de texto foram analisadas, com foco na materialidade textual. A análise revela que as obras de Halliday e Hasan (1976) e Beaugrande e Dressler (1981) são as mais influentes. As definições e caracterizações de texto se distinguem pelo peso dado aos fatores do contexto e da materialidade textual. Enfatiza-se a distinção crucial entre texto e materialidade textual, para que estudos focados na última deixem claras as características específicas do tipo de materialidade investigada em contraste com aquelas generalizáveis a qualquer texto.

Palavras-chave: Texto. Textualidade. Materialidade textual.

* Professor Auxiliar. Estudante de Doutorado no Programa de Pós-Graduação em Difusão do Conhecimento (UFBA, UNEB, IFBA, UEFS, LNCC, SENAI CIMATEC). ORCID: <https://orcid.org/0000-0001-6834-0579>. E-mail: davioliveira@uneb.br.

** Professora Titular, atuando no Programa de Pós-Graduação em Inglês: Estudos Linguísticos e Literários. Doutora em Letras (UFSC). ORCID: <https://orcid.org/0000-0003-4183-8072>. E-mail: leda@cce.ufsc.br.

*** Professor Pleno. Docente do Programa de Pós-Graduação em Modelagem Computacional e Tecnologia Industrial (SENAI CIMATEC) e do Programa de Pós-Graduação em Difusão do Conhecimento (UFBA, UNEB, IFBA, UEFS, LNCC, SENAI CIMATEC). ORCID: <https://orcid.org/0000-0001-7476-9267>. E-mail: hbbpereira@gmail.com.

Abstract: This study investigates answers to the complex question “what is a text?” and proposes an operational definition of text that synthesizes the general characteristics found and specifies some characteristics of textual materiality. Through a systematic review of scientific articles, various definitions and characterizations of text were analyzed, with a focus on textual materiality. The analysis reveals that the works of Halliday and Hasan (1976) and Beaugrande and Dressler (1981) are the most influential. The definitions and characterizations of text are distinguished by the weight given to contextual factors and textual materiality. The crucial distinction between text and textual materiality is emphasized, so that studies focused on the latter make clear the specific characteristics of the type of materiality investigated in contrast to those that can be generalized to any text.

Keywords: Text. Textuality. Textual Materiality.

Resumen: Este estudio investiga respuestas a la compleja pregunta “¿qué es el texto?” y propone una definición operacional de texto que sintetiza las características generales encontradas y especifica algunas características de la materialidad textual. A través de una revisión sistemática de artículos científicos, se analizaron diversas definiciones y caracterizaciones del texto, con foco en la materialidad textual. El análisis revela que las obras de Halliday y Hasan (1976) y Beaugrande y Dressler (1981) son las más influyentes. Las definiciones y caracterizaciones de texto se distinguen por el peso dado a los factores contextuales y a la materialidad textual. Se enfatiza la distinción crucial entre texto y materialidad textual, para que los estudios centrados en esta última dejen claras las características específicas del tipo de materialidad investigada en contraste con aquellas generalizables a cualquier texto.

Palabras clave: Texto. Textualidad. Materialidad textual.

1 INTRODUÇÃO

Respostas à pergunta “o que é um texto?” são de interesse de várias áreas de investigação, encontrando lugar desde a Filosofia, a Hermenêutica e a Crítica Literária (Wilson, 2012), passando pela Educação Matemática (Otte, 1986) e encontrando posição central nas discussões da Análise do Discurso e da Linguística Textual (Titscher *et al.*, 2000) – sem, todavia, convergir para uma única resposta que seja igualmente aceita em todos os casos. Partimos do pressuposto de que parte da falta de consenso é resultado da indissociabilidade de duas facetas do texto enquanto objeto de estudo: sua textualidade e sua materialidade. Por textualidade, referimo-nos às características que fazem um texto ser um texto (Beaugrande; Dressler, 1981) e que são comumente dependentes do contexto. Por materialidade, referimo-nos à realidade física que um texto assume, ou seja, a sequência de símbolos que forma sua superfície linguística, comumente registrados ou registráveis na forma de caracteres impressos e invariáveis em diferentes contextos.

Apesar de não serem suficientes para abarcar todos os fenômenos envolvidos na textualidade, análises mais próximas da materialidade textual, de abordagem quantitativa ou qualiquantitativa, comumente utilizando métodos de Processamento de Linguagem Natural, se têm demonstrado frutíferas para a descoberta de tendências universais, a exemplo dos trabalhos que modelam textos usando redes semânticas ou textuais (Budell *et al.*, 2023; Grabska-Gradzińska *et al.*, 2012; Oliveira; Senna; Pereira, 2024; Stanisz; Drożdż; Kwapieliński, 2024; Teixeira *et al.*, 2010). Tais descobertas são interpretadas como resultado de fatores subjacentes, seja da textualidade, da linguagem ou da cognição humana, e são complementares às análises do texto com foco em suas características contextuais.

A coesão textual, às vezes considerada característica central da textualidade (Halliday; Hasan, 1976), se tem mostrado passível de análises dessa natureza, com alto nível de implementabilidade computacional (Hoey, 1991; Oliveira; Senna; Pereira, 2024). Beaugrande e Dressler (1981, p. 14, tradução nossa) caracterizam coesão como o critério de textualidade que “[...] concerne às formas em que os componentes da superfície textual, *i.e.*, as palavras que ouvimos ou vemos, são mutualmente conectadas em uma sequência”¹ (p. 3). Algo semelhante é proposto por Hoey (1991, p. 3, tradução nossa), que afirma que a “coesão pode ser vagamente definida como a forma com que certas palavras ou características gramaticais de uma sentença podem conectar tal sentença com suas predecessoras (e sucessoras) em um texto”². Assim, dentre as características da textualidade, a coesão é aquela que mais tende a se manter estável em diferentes contextos de uso, assumindo papel central no estudo da materialidade textual.

Pelo acima exposto, o presente estudo busca contribuir com as discussões sobre o texto, enfatizando características da materialidade textual. Para isso, realizamos uma revisão sistemática (Petticrew; Roberts, 2006) para analisar as definições ou caracterizações de texto presentes em trabalhos relacionados à Linguística Textual que mencionem os termos “textualidade” e “coesão”. Inicialmente, analisamos as definições qualitativamente, identificando suas principais características, e, a partir dos resultados dessa análise, realizamos uma análise quantitativa de agrupamento dos estudos selecionados. Com isso, buscamos (a) identificar as definições prevalentes nos estudos em questão; (b) agrupar os estudos a partir das características das definições encontradas; e (c) propor uma definição operacional de texto que contemple as principais características das definições encontradas e que dê conta de características da materialidade textual, contribuindo para análises qualiquantitativas. A definição é operacional por enfatizar características importantes para uma formalização do texto que facilite seu processamento com auxílio de algoritmos computacionais.

Tendo em vista esses objetivos, este artigo foi organizado em duas seções básicas. A segunda descreve o método da revisão sistemática e da análise das definições, de agrupamento e da rede semântica; a terceira apresenta os resultados e a discussão, compreendendo a análise das definições e uma definição operacional de texto.

2 MÉTODO

Esta revisão sistemática foi metodologicamente inspirada no trabalho reportado em Pereira *et al.* (2022a) e, assim como nesse estudo, seguiu as etapas descritas por Petticrew e Roberts (2006): (a) definição da pergunta de pesquisa, (b) definição dos tipos de estudo considerados, (c) busca pelos estudos, (d) filtragem dos estudos, (e) análise dos estudos, (f) síntese dos estudos, e (g) disseminação dos resultados.

Na primeira etapa, buscou-se compreender *como texto é definido ou caracterizado em trabalhos interessados em fatores da textualidade*. Na segunda etapa, considerando a multiplicidade de abordagens de estudo do texto e da textualidade, e que a comparação

¹ No original: “[...] concerns the ways in which the components of the surface text, *i.e.* the actual words we hear or see, are mutually connected within a sequence.”

² No original: “Cohesion may be crudely defined as the way certain words or grammatical features of a sentence can connect that sentence to its predecessors (and successors) in a text.”

exaustiva destas múltiplas abordagens está fora do escopo do presente estudo, nós nos limitamos a investigar os estudos que trouxessem alguma definição ou caracterização de texto, que mencionassem os termos “textualidade” e “coesão” e que tivessem sido publicados como artigos em periódicos revisados por pares e escritos em língua inglesa, portuguesa ou espanhola. Na terceira etapa foram realizadas buscas no ScienceDirect³, Portal de Periódicos da Capes⁴ e no SciELO⁵, por textos contendo a sequência lexical “*a text is a*”, que mencionassem os termos “*textuality*” ou “*text linguistics*” e que mencionassem o termo “*cohesion*”. Assim, a sequência lexical de busca utilizada foi “*a text is a*” AND (“*textuality*” OR “*text linguistics*”) AND “*cohesion*”. Utilizamos termos em inglês na sequência lexical porque textos publicados em outras línguas costumam trazer *abstracts* em inglês. As buscas foram realizadas no dia 11 de abril de 2024.

Na quarta etapa, filtragem dos estudos, consideramos os seguintes critérios de exclusão:

Primeiro Filtro:

1. a versão digital do documento não está disponível abertamente ou por acesso institucional via Universidade do Estado da Bahia (UNEB) ou Instituto Federal de Educação, Ciência e Tecnologia da Bahia (IFBA);
2. a publicação é um capítulo de livro;
3. a publicação é um trabalho publicado em anais de evento;
4. o documento é um trabalho não publicado ou não revisado por pares;
5. a publicação é um relatório técnico;
6. a publicação é uma resenha;
7. a publicação é uma resposta direta a outra publicação;
8. a publicação é uma bibliografia anotada;
9. a publicação é um trabalho de conclusão de curso de graduação ou especialização;
10. a publicação é uma dissertação de mestrado;
11. a publicação é uma tese de doutorado;
12. a publicação não está escrita em inglês, português ou espanhol.

Segundo Filtro:

1. a publicação não traz uma definição ou caracterização explícita do conceito de texto.

A Figura 1 mostra o resultado das buscas e das filtrações. Vale observar que, mesmo com o foco em estudos mencionando “textualidade” e contendo definições ou caracterizações de texto, 75% dos estudos não passaram pelo segundo filtro, evidenciando uma ausência destas definições ou caracterizações na maioria dos estudos. Este resultado evidencia o paradoxo notado por Wilson (2012, p. 342, tradução nossa) de que “‘o texto’ é, ao mesmo tempo, o mais teorizado e o menos teorizado dentre os conceitos”⁶.

³ Disponível em: <https://www.sciencedirect.com/>.

⁴ Disponível em: <https://www.periodicos.capes.gov.br/>.

⁵ Disponível em: <https://scielo.org/>.

⁶ No original: “‘the text’ is at once the most theorised and least theorised of concepts.”

Busca	"a text is a" AND ("textuality" OR "text linguistics") AND "cohesion"		
Plataforma	ScienceDirect	Periódicos Capes	SciELO
Resultados	210	49	2
Únicos	210	38	2
Primeiro filtro	179	15	2
Segundo filtro	36	11	2
Selecionados	49		
Duplicados	1		
Analizados	48		

Figura 1 – Resultado das buscas

Fonte: Elaborada pelos autores

A Figura 2, por sua vez, mostra a quantidade de publicações por década, evidenciando uma tendência de crescimento das publicações disponibilizadas em meio digital, além de resumir a amplitude do período contemplado – da década de 1970 à década de 2020 – e a distribuição dos anos de publicação, com aproximadamente metade dos estudos selecionados publicada entre as décadas de 1970 e 2010 e a outra metade, aproximadamente, publicada entre 2010 e 2020.

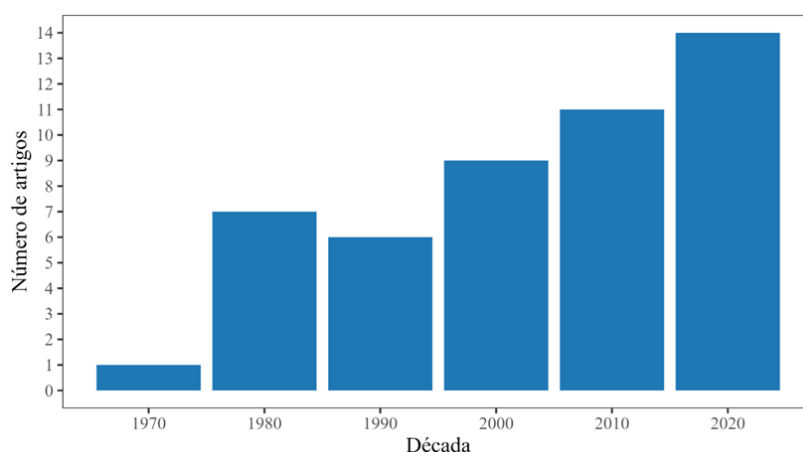


Figura 2 – Número de artigos selecionados por década

Fonte: Elaborada pelos autores.

Na quinta etapa, para analisar as definições e caracterizações de texto encontradas, selecionamos primeiramente as citações em que estas aparecem explicitamente. Em seguida, destacamos as referências mencionadas nas citações ou nos mesmos parágrafos das citações, e que foram utilizadas como base para a definição, como citação direta ou indireta. Por fim, criamos paráfrases parciais para cada citação, mantendo apenas as informações diretamente relacionadas à definição de texto e fazendo ajustes necessários (as citações e paráfrases estão disponíveis em Oliveira, 2024). Estes ajustes incluíram a remoção de aspas, referências e páginas de citações; a tradução para o inglês, no caso de citações em português e espanhol; a mudança de algumas palavras para a grafia do padrão norte-americano (por exemplo, “*rigourously*” para “*rigorously*”); a remoção de afirmações negativas sobre a definição de texto, já que elas funcionam mais como explicações sobre o construto do que como uma definição propriamente dita (por exemplo, “*The text is not a display of language resources [...]*”). A quinta etapa foi dividida em três fases de análise: (a) análise dos trabalhos mais referenciados nas definições; (b) análise de agrupamento com base nas características das definições; (c) análise das paráfrases das citações.

Para a análise dos trabalhos mais referenciados identificamos primeiramente quais deles eram diferentes edições ou republicações do mesmo texto – que foram considerados como o mesmo trabalho. Em seguida, contabilizamos quantos dos artigos selecionados mencionaram cada um dos trabalhos.

Para a análise do agrupamento, fizemos a leitura e a análise qualitativa das características encontradas nas definições. A análise qualitativa consistiu na identificação de categorias representadas por afirmações que fossem verdadeiras para pelo menos uma definição, por exemplo: “a definição menciona coesão ou que o texto funciona como uma unidade”. Depois da identificação das categorias, categorizamos cada trabalho. Para isto, para cada estudo foi atribuído o valor 1 ou 0 em cada uma das categorias: 1 no caso de pelo menos uma das definições trazidas pertencer à categoria, ou 0 caso contrário. A partir desses valores, foram calculadas as distâncias euclidianas entre os trabalhos e foi feita uma análise de agrupamento hierárquico utilizando o método Ward (Murtagh; Legendre, 2014).

As paráfrases das citações contendo definições ou caracterizações de texto foram elaboradas conforme procedimento explicado anteriormente, para a última parte da análise que foi baseada na criação de uma rede semântica (dados disponíveis em Oliveira, 2024). Uma rede é uma estrutura formada por pontos, chamados vértices, e linhas representando as relações entre os pontos, chamadas arestas (Newman, 2010). Redes semânticas, também chamadas redes textuais (Segev, 2020), são aquelas cujos vértices representam unidades textuais, como palavras, lemas ou radicais, e cujas arestas representam uma relação semântica, sintática ou de coocorrência entre essas unidades, constituindo sistemas imbuídos com intenção de funcionalidade e estabelecidos por um contexto (Grilo *et al.*, 2017). Para uma discussão detalhada sobre outras definições existentes na literatura, nos referimos a Pereira *et al.* (2022a). Como a rede construída não distingue relações lexicais afirmativas e negativas, mantivemos apenas as afirmativas para a rede representar um delineamento afirmativo da definição de texto a partir dos artigos analisados.

Para a construção da rede, as palavras das paráfrases de definição ou caracterização de texto foram lematizadas, ou seja, tiveram suas flexões removidas, mantendo-se apenas o lema, ou forma canônica. Em seguida, as paráfrases foram divididas em períodos, considerando o caractere “.” (ponto) como delimitador. Então, geramos redes em que cada período é representado por um conjunto de vértices completamente interconectados, estrutura chamada de clique, formando uma rede de cliques (Fadigas; Pereira, 2013). Assim, os vértices representam os lemas únicos do texto enquanto as arestas representam a coocorrência em um mesmo período na rede analisada. Além disso, palavras funcionais com alta frequência (por exemplo, artigos e preposições) foram removidas. Consideramos como vértices mais importantes da rede aqueles identificados como polos (*hubs*) seguindo a definição de Silva *et al.* (2012, p. 7) que consideram como polo cada vértice cujo grau seja maior ou igual a $\langle k \rangle + 2\sigma$, sendo $\langle k \rangle$ o grau médio e σ o desvio padrão da distribuição de graus da rede.⁷ Outros estudos em diferentes áreas aplicaram exitosamente este critério evidenciando sua robustez na identificação de polos, a exemplo de Rosário *et al.* (2015) e Pereira *et al.* (2022b).

3 RESULTADOS E DISCUSSÃO

Nesta seção, apresentamos a análise das definições e caracterizações de texto encontradas, juntamente com uma proposta de definição operacional. Primeiramente, na subseção 3.1, examinamos as definições mais recorrentes nos trabalhos selecionados, destacando a influência do trabalho de Halliday e Hassan (1976). Em seguida, a subseção 3.2 expõe os resultados da análise de agrupamento das definições. Posteriormente, a subseção 3.3 apresenta os resultados da análise da rede semântica construída a partir das paráfrases das definições. Finalmente, na subseção 3.4, sintetizamos as definições discutidas e propomos uma definição operacional que considera a unidade semântica do texto, com ênfase nas características de sua materialidade textual.

3.1 ANÁLISE DAS DEFINIÇÕES REFERENCIADAS NOS TRABALHOS

A definição de texto de Halliday e Hasan (1976) é a mais recorrente no *corpus*, aparecendo em 10 dos 48 artigos examinados (Albarrán-Santiago, 2015; Blom; Hansen, 2015; Carrell, 1982; Cheung, 2010; Dalalyan, 2015; Kaplan; Grabe, 2002; Khalil, 1989; Kitis; Milapides, 1997; Marashi, 2020; Yang; Sun, 2012).

O segundo lugar é dividido por dois trabalhos: um é outro livro dos mesmos autores (Halliday; Hasan, 1989, ou sua primeira edição de 1985), mencionado em 4 artigos (Fang, 1997; Golden; Pappas, 1990; Kaplan; Grabe, 2002; Purser; Dreyfus; Jones, 2020), o outro é o de Beaugrande e Dressler (1981, a reimpressão de 1988 ou a edição em espanhol de 1997), mencionado em 4 artigos (Dilidüzgün, 2013; Lyons, 2021; Paz Campuzano; Serva Paucar, 2022; Rottensteiner, 2010).

⁷ Os artigos selecionados, as definições, referências, paráfrases e o código utilizado na construção das redes e na análise dos dados podem ser conferidos em Oliveira (2024).

A famosa definição de Halliday e Hasan (1976, p. 1) apresenta o texto como “[...] uma unidade de linguagem em uso”, enfatizando que se trata de uma unidade semântica, e não formal⁸, e dá grande ênfase ao contexto. A definição de texto de Beaugrande e Dressler (1981) diferencia-se da de Halliday e Hasan (1976) por trazer sete critérios de textualidade: (a) coesão, a interconexão de elementos da superfície do texto; (b) coerência, a configuração semântica do texto considerando a interconexão de seus significados; (c) intencionalidade, as atitudes de quem produz o texto buscando atingir níveis aceitáveis de coesão e coerência; (d) aceitabilidade, as atitudes de quem lê o texto pressupondo níveis aceitáveis de coesão e coerência; (e) informatividade, a diferença entre o que é conhecido e o que não é conhecido por quem produz ou lê o texto em uma dada situação comunicativa, (f) situacionalidade, fatores da situação de uso do texto que o tornam relevante em tal situação; e (g) intertextualidade, a conexão e relação de dependência entre diferentes textos (estes critérios, exceto informatividade, são apresentados inicialmente em Beaugrande, 1980; os sete critérios são detalhados em Beaugrande; Dressler, 1981).

Quatro definições dividem o terceiro lugar, e são referenciadas em dois artigos cada. A definição de Halliday (1970) é mencionada por Cheung (2010) e Makino (1982); a definição de Halliday (1977) é mencionada por Dalalyan (2015) e Dilidüzgün (2013); a definição de Brown e Yule (Brown; Yule, 1983, ou sua reimpressão de 1984), é mencionada por Amaral (1985) e Cheung (2010); a definição de Günay (2003, ou sua edição de 2007), por fim, é mencionada por Bayat e Yurdakul (2014) e Ocak e Baysal (2016).

Além dos já mencionados trabalhos de Halliday, outros 4 trabalhos do mesmo autor são mencionados (uma vez cada), evidenciando sua influência nas discussões sobre texto e textualidade. Esses trabalhos contêm definições semelhantes à encontrada em Halliday e Hasan (1976). Brown e Yule (1983, p. 190, tradução nossa) apresentam texto como “[...] um registro verbal de um evento comunicativo”⁹ e, logo em seguida, referenciam Halliday e Hasan (1976) para definir texto de forma mais precisa. A definição de Günay enfatiza o texto enquanto sistema cujos elementos interconectados são produzidos em forma oral ou escrita. Os demais trabalhos foram referenciados em apenas um artigo.

Em síntese, as definições de texto propostas por Halliday e Hasan predominam no *corpus*, seguidas pela definição de Beaugrande e Dressler, frequentemente usada para complementá-las.

Tanto as definições de Halliday e Hasan (1976) quanto a de Beaugrande e Dressler (1981) apontam para uma característica importante de textos: seus significados e funcionalidade são fortemente dependentes do contexto, que inclui seus produtores e leitores. Se, por um lado, essa característica aponta para a importância de considerar fatores contextuais no estudo do texto, por outro aponta para a complexidade de uma teoria que dê conta do maior número possível desses fatores. Uma abordagem interdisciplinar é uma possibilidade de solução. Enquanto áreas como a Análise do Discurso conseguem dar conta de fenômenos mais vinculados ao contexto, áreas como a

⁸ No original: “A text is best regarded as a SEMANTIC unit: a unit not of form but of meaning” (Halliday; Hasan, 1976, p. 2)

⁹ No original: “[...] a verbal record of a communicative event”.

Linguística Quantitativa e a Física dão conta de fenômenos mais vinculados à materialidade textual. Unir estes resultados pode possibilitar uma visão mais completa da complexidade do texto. A definição operacional que propomos na seção 3.4 é um passo em direção a essa unificação.

3.2 ANÁLISE DE AGRUPAMENTOS BASEADOS NAS DEFINIÇÕES

A partir da análise qualitativa das definições e caracterizações de texto, foram identificadas 22 características. Estas características consideram a relevância dada a cada um dos seguintes atributos do texto para sua definição, apresentados aqui em ordem decrescente de frequência: (a) coesão, ou funcionamento como unidade; (b) nível lógico ou semântico; (c) função comunicativa; (d) situacionalidade, ou situação ou contexto de uso; (e) aceitabilidade, ou as atitudes do leitor ou ouvinte em receber o texto como tal; (f) modalidade escrita; (g) modalidade oral; (h) intencionalidade, ou as atitudes de seu produtor; (i) unidade formal ou estrutural; (j) coerência; (k) nível discursivo; (l) nível lexical; (m) irrelevância do tamanho em sua definição; (n) caráter linear ou sequencial; (o) informatividade; (p) intertextualidade; (q) nível sintático; (r) distinção de um conjunto aleatório de sentenças; (s) nível fonológico ou prosódico; (t) estilo; (u) padrões de repetição ou similaridade de elementos; e (v) possibilidade de ser não-verbal.

As características identificadas foram utilizadas na análise de agrupamento hierárquico. A Figura 3, mais adiante, apresenta o dendrograma gerado a partir da análise, destacando a divisão em três grupos. As características identificadas em cada estudo estão disponíveis no Apêndice A e o Apêndice B apresenta visualmente, com o uso das cores da Figura 3, a distribuição das características em cada um dos grupos.

Consideramos como características predominantes de cada grupo aquelas presentes em pelo menos metade dos trabalhos do grupo e ausentes na maioria dos artigos dos outros grupos. O Grupo 1 se diferencia dos demais por privilegiar em suas definições a coesão, a função comunicativa do texto, a situacionalidade e a aceitabilidade. Além disso, essas definições registram, mesmo que em menor frequência, a importância da intencionalidade, da intertextualidade e do nível lógico ou semântico do texto. Assim, as definições deste grupo definem o texto como unidade semântica, com função comunicativa, dependente do contexto, influenciado por seus processos de produção e uso, e interconectado com outros textos.

O Grupo 2 se diferencia dos demais principalmente pela sua heterogeneidade e por não apresentar características predominantes. Independentemente, algumas características frequentes são a importância da coesão, o nível lógico ou semântico e a possibilidade de o texto ser definido por unidade formal ou estrutural. Esta última característica merece destaque por ser contrária à ênfase dada por Halliday e Hasan de que texto não é definido por sua unidade formal ou estrutural, mas por sua unidade semântica. No total, 13 dos 48 artigos selecionados trazem a unidade formal ou estrutural na definição de texto, sendo 6 do Grupo 1 (Dilidüzgün, 2013; Fowler, 1979; Luke, 1989; Roman; Roman, 2014; Rottensteiner, 2010; Yang; Sun, 2012), 5 do Grupo 2

(Beaugrande, 1979, 1980; Okurowski, 1989; Oliveira; Senna; Pereira, 2024; Sándor, 1989) e 2 do Grupo 3 (Bayat; Yurdakul, 2014; Felisbino, 2001).

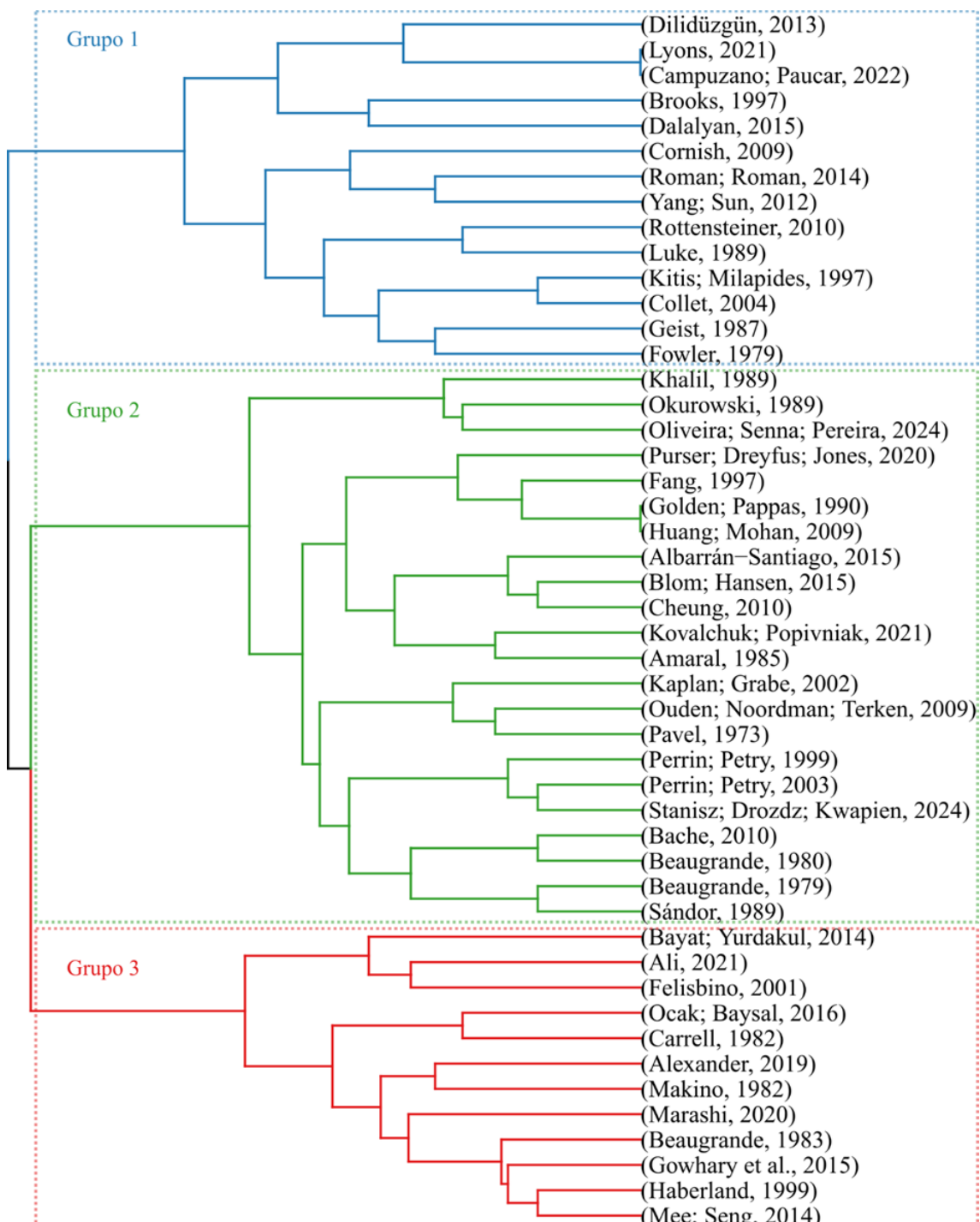


Figura 3 – Dendrograma da análise de agrupamento dos trabalhos

Fonte: Elaborada pelos autores.

Por fim, o Grupo 3 se diferencia dos demais por dar ênfase à possibilidade de o texto se apresentar na modalidade escrita ou oral. Esta distinção é importante porque alguns autores (Kovalchuk; Popivniak, 2021; Rottensteiner, 2010) definem texto de forma limitada ou diretamente dependente da modalidade escrita. Apesar das vantagens da modalidade escrita para a análise de textos, principalmente com foco na materialidade textual, uma definição completa de texto deve incluir textos orais. Além disso, considerando o texto enquanto unidade semântica e considerando que atribuímos sentido não apenas a palavras como também a gestos, expressões faciais e outros sistemas de significação não-verbais, uma definição completa de texto deve incluir também textos não-verbais, característica considerada apenas em um artigo (Cornish, 2009). Embora reconheçamos a importância do estudo de textos não-verbais, o status de alguns desses artefatos como texto ainda é uma questão em aberto. Como exemplificaremos na seção 3.4, a definição que propomos abrange a materialidade de um subconjunto desses artefatos comumente considerados textos, como produções em língua de sinais e escrita pictórica, mas exclui casos que merecem atenção em estudos futuros, como cartuns, fotografias e pinturas.

Uma maneira de conciliar as definições que consideram um texto como unidade formal (com foco na materialidade textual) com aquelas que consideram o texto como uma unidade semântica (com foco em fatores contextuais da textualidade) é entender a unidade formal do texto como uma consequência da sua unidade semântica. A ausência de unidade formal ou estrutural não implica necessariamente a ausência de unidade semântica: a presença de uma unidade formal é um indicativo de unidade semântica subjacente.

3.3 ANÁLISE DE REDE SEMÂNTICA DAS PARÁFRASES DAS DEFINIÇÕES

A Figura 4 ilustra a rede semântica construída a partir das paráfrases das definições e caracterizações de texto (disponíveis em Oliveira, 2024). Os vértices considerados polos da rede, apresentados em ordem decrescente considerando o grau, foram os que representam os seguintes lemas: ‘*text*’, ‘*discourse*’, ‘*context*’, ‘*system*’, ‘*language*’, ‘*cohesion*’, ‘*unit*’, ‘*linguistic*’, ‘*structure*’, ‘*coherence*’, ‘*semantic*’.

A análise da rede não apenas corrobora os resultados da análise de agrupamentos, mas também a complementa ao evidenciar que o conceito de discurso é fortemente relacionado ao de texto. Um problema, porém, é a polissemia da palavra *discurso*, muitas vezes utilizado como sinônimo, principalmente em língua inglesa. A solução proposta é considerar que, quando utilizado como sinônimo de texto, discurso pode ser representado pelo lema “*texto*”, e quando utilizado com outro significado, pode ser considerado como parte do contexto de produção ou recepção do texto. Assim, na próxima seção propomos uma definição operacional de texto que sintetize as definições encontradas e que dê conta de características da materialidade textual, buscando utilizar os lemas identificados como polos da rede, exceto ‘*discurso*’.

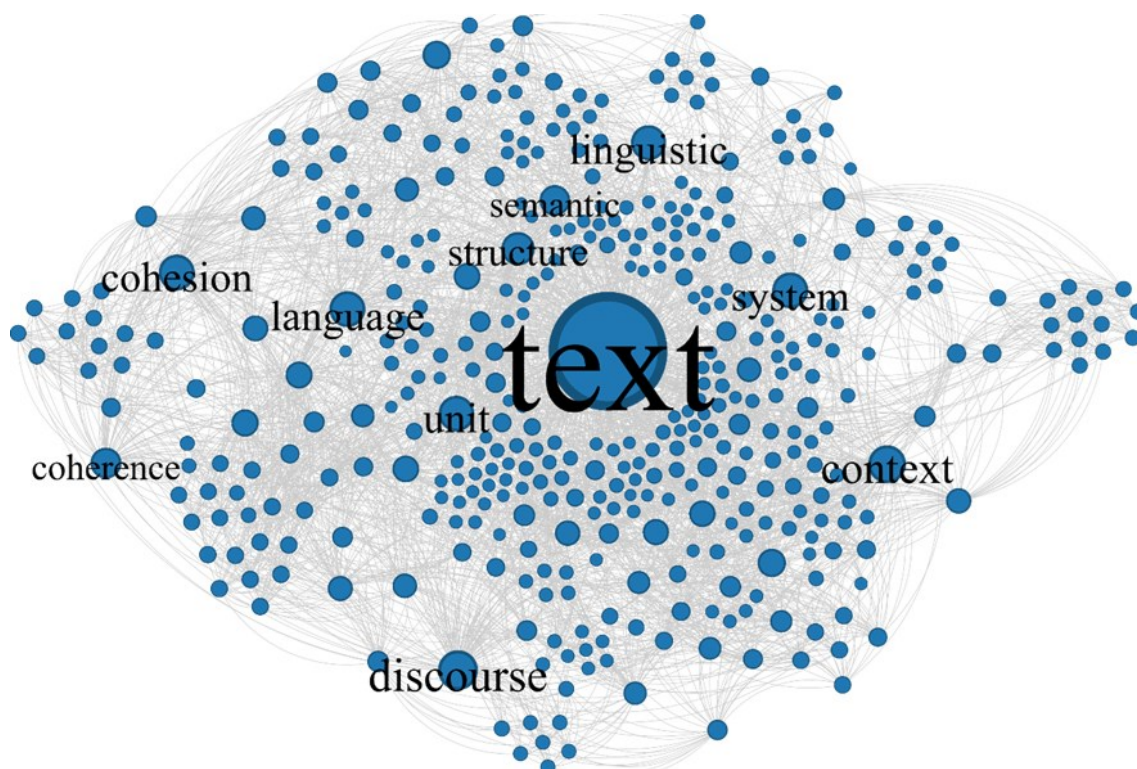


Figura 4 – Rede semântica das paráfrases das definições encontradas

Nota: Vértices representam os lemas únicos das paráfrases das definições e caracterizações de texto encontradas. Arestas representam a coocorrência em um mesmo período. O tamanho dos vértices representa o grau. Os vértices acompanhados pela palavra que representa são aqueles identificados como polos (*hubs*) da rede.

Fonte: Elaborada pelos autores.

3.4 PROPOSIÇÃO DE UMA DEFINIÇÃO OPERACIONAL DE TEXTO COM FOCO NA MATERIALIDADE

Como dito anteriormente, o texto é comumente caracterizado na literatura investigada por sua unidade semântica ou por sua unidade formal. Defendemos que comumente os dois aspectos estão conectados. Ou seja, mesmo que um prescinda do outro, a tendência é que uma unidade formal reflita uma unidade semântica subjacente.

Posto isso – apesar de reconhecermos as limitações deste estudo – propomos, a seguir, uma **definição operacional de texto**, que busca contemplar a materialidade textual, permitindo-nos formalizá-lo e representá-lo quantitativamente, de forma compatível com definições focadas na textualidade:

Um texto é uma unidade semântica, cujos significados são fortemente influenciados pelos contextos de produção e recepção, materializada através de um sistema de linguagem. Sua materialidade é tipicamente uma estrutura coesa. Considerando especificamente sistemas linguísticos, um texto é tipicamente materializado por uma sequência de segmentos $\mathcal{T} = (\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_i, \dots, \mathcal{S}_n)$, sendo cada segmento $\mathcal{S}_i$ uma sequência $\mathcal{S} = (t_1, t_2, \dots, t_j, \dots, t_n)$ de tokens t_j , sendo cada token, por sua vez, um conjunto de unidades linguísticas mínimas.

A definição operacional proposta é semelhante à definição de discurso de Chan (2004, p. 141; 2006, p. 761) ao considerar um texto como uma sequência de segmentos. É diferente, porém, por se propor a definir texto e não discurso e por trazer uma definição dos segmentos que compõem o texto na superfície linguística, ou seja, em sua materialidade, como conjunto de unidades linguísticas mínimas. A principal implicação da definição proposta é que, para ser um texto, o texto tem de ser materializado, já que um “texto” não-materializado não é passível de análise.

Vale notar que essa definição operacional pode ser ajustada para comportar outros níveis da materialidade textual, por exemplo, considerando cada segmento como um parágrafo formado por uma sequência de períodos que, por sua vez, são formados por uma sequência de tokens. Também vale ressaltar que utilizamos o termo *linguagem* de forma vaga, de modo semelhante ao proposto por Saussure (2006, p. 15–25) e diferenciando-se do conceito de língua, limitado à linguagem verbal. Além disso, o conceito de linguagem não se limita à competência linguística, *I-language*, como definida por Chomsky (2015). Como nos interessamos pela materialidade textual, também estamos interessados nas características do uso da linguagem, ou para usar outro termo proposto por Chomsky (2015), nas características da performance.

A principal característica do sistema de linguagem a que a definição operacional proposta se refere é a de ele ser formado por partes que possam ser consideradas suas unidades mínimas. Assim, tais unidades são grafemas ou caracteres na análise de textos escritos, fonemas na análise de textos orais, ou queremas na análise de textos em línguas de sinais. Apesar de não ser uma lista exaustiva, estes três tipos constituem um ponto de partida para a análise da materialidade textual por darem conta das modalidades orais, escritas e gestuais das diferentes línguas e das línguas de sinais.

Como mencionado anteriormente, a definição consegue abranger um subconjunto de textos não-verbais, mas exclui casos que merecem atenção. A definição proposta pode ser dividida em duas partes: uma centrada na textualidade (“Um texto é uma unidade semântica [...] tipicamente uma estrutura coesa.”) e outra focada na materialidade textual (“Considerando especificamente sistemas linguísticos [...] um conjunto de unidades linguísticas mínimas.”). Tomando uma pintura como exemplo, a aplicação da primeira parte da definição não é problemática. Porém, atribuir a uma pintura uma estrutura formada por segmentos sequenciais de unidades menores também sequenciais, se for possível, não é trivial. Assim, defendemos que se um artefato puder ser descrito pela primeira parte da definição, pode ser considerado um texto. Se também puder ser descrito pela segunda parte, trata-se de um texto que pode ser submetido a análises compatíveis com esse tipo de materialidade, como representação em redes, análise de coesão baseada na repetição de unidades (Oliveira; Senna; Pereira, 2024) e transcrição automatizada para a modalidade verbal. Os casos que não se encaixam na segunda parte são textos, ou textos-em-potencial, cuja materialidade ainda precisa ser formalmente definida para possibilitar esse mesmo tipo de análise.

Apesar de a definição de linguagem e até a definição de ‘unidade semântica’, conceitos importantes para a definição, serem pontos de discussão, consideramos que a definição operacional de texto aqui proposta sintetiza as definições encontradas na literatura revisada dando conta de características da materialidade textual e sendo útil para investigações sobre *texto* com este foco.

4 CONSIDERAÇÕES FINAIS

Embora seja difícil responder à pergunta “o que é texto?” ou “o que caracteriza um texto?” de forma categórica, suas possíveis respostas são o ponto de partida para investigações em várias áreas de estudo que têm o texto como objeto de estudo central, e um ponto de interesse para várias outras que se ocupam de suas propriedades enquanto sistema de informação e representação do conhecimento. Este estudo buscou contribuir com as discussões sobre texto e textualidade a partir da análise de definições e caracterizações encontradas na literatura, com base em uma revisão sistemática. Dentro desse contexto, destacamos a proposição de uma definição operacional de texto.

Conforme os resultados obtidos, defendemos a distinção entre texto e materialidade textual, considerando esta última um aspecto do primeiro. Essa distinção se fundamenta em dois pontos principais:

- a) algumas características da materialidade textual podem ser específicas do sistema de linguagem ou modalidade em questão, o que torna crucial a interpretação dos resultados à luz dessa especificidade;
- b) a identificação de padrões na materialidade textual presentes em diferentes modalidades e sistemas de linguagem deve ser central nas discussões sobre textualidade.

Elucidar essa distinção entre texto e materialidade textual permite maior precisão na identificação de características centrais à textualidade, presentes em diversas modalidades de texto, e na diferenciação destas em face de características marginais, restritas a algumas modalidades. Ao propor essa distinção, este artigo contribui para uma possível convergência entre estudos do texto em diferentes disciplinas. Ressaltamos que mesmo estudos limitados à superfície linguística – como muitos na Linguística Quantitativa e na Física – podem contribuir para a compreensão da textualidade ao identificarem características em diversas modalidades de materialidade.

A revisão apresentada neste trabalho limitou-se a estudos que mencionaram o termo “coesão” em vista do interesse em definições que dessem conta da materialidade textual. Estudos futuros podem complementar os resultados apresentados a partir da revisão de trabalhos com outros enfoques, como em textos não-verbais, de forma a ampliar a compreensão sobre *texto*, objeto multifacetado e complexo.

REFERÊNCIAS

- ALBARRÁN-SANTIAGO, M. La lingüística textual en los trabajos de investigación producidos por alumnado universitario: enseñanza en las planificaciones de aula y los libros de texto. *Revista Electrónica Educare*, v. 19, n. 1, p. 389-406, 1 jan. 2015.
- AMARAL, M. P. As categorias de coesão textual e a complexidade do texto. *Letras de Hoje*, v. 20, n. 2, p. 29-40, 1985.
- BAYAT, N.; YURDAKUL, Y. The degree of textuality in speech of six-year-old children. *Procedia - Social and Behavioral Sciences*, v. 116, p. 3874-3879, fev. 2014. DOI: <https://doi.org/10.1016/j.sbspro.2014.01.858>.

- BEAUGRANDE, R.-A. Toward a general theory of creativity. *Poetics*, v. 8, n. 3, p. 269-306, jun. 1979. DOI: [https://doi.org/10.1016/0304-422X\(79\)90036-6](https://doi.org/10.1016/0304-422X(79)90036-6).
- BEAUGRANDE, R.-A. De. The pragmatics of discourse planning. *Journal of Pragmatics*, v. 4, p. 15-42, 1980. DOI: [https://doi.org/10.1016/0378-2166\(80\)90015-6](https://doi.org/10.1016/0378-2166(80)90015-6).
- BEAUGRANDE, R.-A. De; DRESSLER, W. *Introduction to Text Linguistics*. London: Longman, 1981.
- BLOM, J. N.; HANSEN, K. R. Click bait: forward-reference as lure in online news headlines. *Journal of Pragmatics*, v. 76, p. 87-100, jan. 2015. DOI: <https://doi.org/10.1016/j.pragma.2014.11.010>.
- BROWN, G.; YULE, G. *Discourse analysis*. Cambridge: Cambridge University Press, 1983.
- BUDEL, G.; JIN, Y.; VAN MIEGHEM, P.; KITSACK, M. Topological properties and organizing principles of semantic networks. *Scientific Reports*, v. 13, n. 1, p. 11728, 20 jul. 2023. DOI: <https://doi.org/10.1038/s41598-023-37294-8>.
- CARRELL, P. L. Cohesion is not coherence. *TESOL Quarterly*, v. 16, n. 4, p. 479, dez. 1982. DOI: <https://doi.org/10.2307/3586466>.
- CHAN, S. W. K. Automatic discourse structure detection using shallow textual continuity. *International Journal of Human-Computer Studies*, v. 61, n. 1, p. 138-164, jul. 2004. DOI: <https://doi.org/10.1016/j.ijhcs.2003.12.002>.
- CHAN, S. W. K. Beyond keyword and cue-phrase matching: a sentence-based abstraction technique for information extraction. *Decision Support Systems*, v. 42, n. 2, p. 759-777, nov. 2006. DOI: <https://doi.org/10.1016/j.dss.2004.11.017>.
- CHEUNG, M. The globalization and localization of persuasive marketing communication: a cross-linguistic socio-cultural analysis. *Journal of Pragmatics*, v. 42, n. 2, p. 354-376, fev. 2010. DOI: <https://doi.org/10.1016/j.pragma.2009.06.012>.
- CHOMSKY, N. *The Minimalist Program*. 20. anniversary ed. Cambridge: The MIT Press, 2015.
- CORNISH, F. Inter-sentential anaphora and coherence relations in discourse: a perfect match. *Language Sciences*, v. 31, n. 5, p. 572-592, set. 2009. DOI: <https://doi.org/10.1016/j.langsci.2008.06.002>.
- DALALYAN, M. Textual repetition as a manifestation of irony. *Armenian Folia Anglistika*, v. 11, n. 2 (14), p. 49-55, 15 out. 2015. DOI: <https://doi.org/10.46991/AFA/2015.11.2.049>.
- DILIDÜZGÜN, Ş. The text-structural activities in ELT and (T)urkish (L)anguage (T)eaching. *Procedia - Social and Behavioral Sciences*, v. 70, p. 1432-1437, jan. 2013. DOI: <https://doi.org/10.1016/j.sbspro.2013.01.207>.
- FADIGAS, I. S.; PEREIRA, H. B. B. A network approach based on cliques. *Physica A: Statistical Mechanics and its Applications*, v. 392, n. 10, p. 2576-2587, 2013. DOI: <https://doi.org/10.1016/j.physa.2013.01.055>.
- FANG, Z. A study of changes and development in children's written discourse potential. *Linguistics and Education*, v. 9, n. 4, p. 341-367, jan. 1997. DOI: [https://doi.org/10.1016/S0898-5898\(97\)90005-X](https://doi.org/10.1016/S0898-5898(97)90005-X).
- FELISBINO, F. Avaliação de produção textual. *Linguagem em (Dis)curso*, v. 1, n. 2, 2001.
- FOWLER, R. Preliminaries to a sociolinguistic theory of literary discourse. *Poetics*, v. 8, n. 6, p. 531-556, dez. 1979. DOI: [https://doi.org/10.1016/0304-422X\(79\)90032-9](https://doi.org/10.1016/0304-422X(79)90032-9).
- GOLDEN, J. M.; PAPPAS, C. C. A sociolinguistic perspective on retelling procedures in research on children's cognitive processing of written text. *Linguistics and Education*, v. 2, n. 1, p. 21-41, jan. 1990. DOI: [https://doi.org/10.1016/S0898-5898\(05\)80009-9](https://doi.org/10.1016/S0898-5898(05)80009-9).
- GRABSKA-GRADZIŃSKA, I.; KULIG, A.; KWAPIEN, J.; DROŽDŽ, S. Complex network analysis of literary and scientific texts. *International Journal of Modern Physics C*, v. 23, n. 07, p. 1250051, jul. 2012. DOI: <https://doi.org/10.1142/S0129183112500519>.
- GRILO, M.; FADIGAS, I. S.; MIRANDA, J. G. V.; CUNHA, M. V.; MONTEIRO, R. L. S.; PEREIRA, H. B. B. Robustness in semantic networks based on cliques. *Physica A: Statistical Mechanics and its Applications*, v. 472, p. 94-102, 2017. DOI: <https://doi.org/10.1016/j.physa.2016.12.087>.
- GÜNAY, V. D. *Metin Bilgisi*. İstanbul: Multilingual Yayınları, 2003.
- HALLIDAY, M. A. K. Language structure and language function. In: LYONS, J. (Ed.). *New Horizons in Linguistics*. Harmondsworth: Penguin Books, 1970. p. 140-165.
- HALLIDAY, M. A. K. Text as semantic choice in social contexts. In: VAN DIJK, T. A.; PETÖFI, J. S. (Eds.). *Grammars and Descriptions*. Research in Text Theory. Berlin: de Gruyter, 1977. p. 176-226.
- HALLIDAY, M. A. K.; HASAN, R. *Cohesion in English*. London: Longman Group Limited, 1976.

- HALLIDAY, M. A. K.; HASAN, R. Language, context, and text: aspects of language in a social-semiotic perspective. 2. ed. Oxford: Oxford University Press, 1989.
- HOEY, M. *Patterns of lexis in text*. Oxford: Oxford University Press, 1991.
- KAPLAN, R. B.; GRABE, W. A modern history of written discourse analysis. *Journal of Second Language Writing*, v. 11, n. 3, p. 191-223, ago. 2002. DOI: [https://doi.org/10.1016/S1060-3743\(02\)00085-1](https://doi.org/10.1016/S1060-3743(02)00085-1).
- KHALIL, A. A study of cohesion and coherence in Arab EFL college students' writing. *System*, v. 17, n. 3, p. 359-371, jan. 1989. DOI: [https://doi.org/10.1016/0346-251X\(89\)90008-0](https://doi.org/10.1016/0346-251X(89)90008-0).
- KITIS, E.; MILAPIDES, M. Read it and believe it: how metaphor constructs ideology in news discourse: a case study. *Journal of Pragmatics*, 1997. DOI: [https://doi.org/10.1016/S0378-2166\(97\)00075-1](https://doi.org/10.1016/S0378-2166(97)00075-1).
- KOVALCHUK, I. V.; POPIVNAK, O. O. Textuality standards of goods labels and packaging. *Current Issues of Ukrainian Linguistics: Theory and Practice*, n. 43, p. 22-41, 2021. DOI: <https://doi.org/10.17721/APULTP.2021.43.22-41>.
- LUKE, A. Open and closed texts: the ideological/semantic analysis of textbook narratives. *Journal of Pragmatics*, v. 13, p. 53-80, 1989. DOI: [https://doi.org/10.1016/0378-2166\(89\)90109-4](https://doi.org/10.1016/0378-2166(89)90109-4).
- LYONS, M. A. Local incoherence, global coherence? Allusion and the readability of ancient Israelite literature. *Old Testament Essays*, v. 34, n. 1, p. 141-164, 2021. DOI: <http://dx.doi.org/10.17159/2312-3621/2021/v34n1a9>.
- MAKINO, S. Japanese grammar and functional grammar. *Lingua*, v. 57, n. 2-4, p. 125-173, jun. 1982. DOI: [https://doi.org/10.1016/0024-3841\(82\)90003-1](https://doi.org/10.1016/0024-3841(82)90003-1).
- MARASHI, S. M. Impact of teaching cohesive devices to Iranian EFL learners: case of IELTS writing task 2 examination. *JL3T (Journal of Linguistics, Literature and Language Teaching)*, v. 6, n. 2, p. 115-126, 2020. DOI: <https://doi.org/10.32505/jl3t.v6i2.2050>.
- MURTAGH, F.; LEGENDRE, P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of Classification*, v. 31, n. 3, p. 274-295, out. 2014. DOI: <https://doi.org/10.1007/s00357-014-9161-z>.
- NEWMAN, M. E. J. *Networks*. Oxford: Oxford University Press, 2010.
- OCAK, G.; BAYSAL, E. A. Evaluation of English textbooks in terms of textuality standards. *Journal of Education and Training Studies*, v. 4, n. 12, p. 36-44, 2016. DOI: <https://doi.org/10.11114/jets.v4i12.1897>.
- OKUROWSKI, M. E. Textual cohesion in Modern Standard Chinese. *Language Sciences*, v. 11, n. 1, p. 89-104, jan. 1989. DOI: [https://doi.org/10.1016/0388-0001\(89\)90016-8](https://doi.org/10.1016/0388-0001(89)90016-8).
- OLIVEIRA, D. A. Revisão sistemática de definições e caracterizações de texto: uma proposta de definição operacional [Dados e código fonte] (v1.0.2). Zenodo, 2024. DOI: <https://doi.org/10.5281/zenodo.11075343>.
- OLIVEIRA, D. A.; SENNA, V.; PEREIRA, H. B. B. Indices of textual cohesion by lexical repetition based on semantic networks of cliques. *Expert Systems with Applications*, v. 237, n. 2024, p. 121580, 2024. DOI: <https://doi.org/10.1016/j.eswa.2023.121580>.
- OTTE, M. What is a text? In: CHRISTIANSEN, B.; HOWSON, A. G.; OTTE, M. (Eds.). *Perspectives on Mathematics Education*. Dordrecht: Springer Netherlands, 1986. p. 173-203.
- PAZ CAMPUZANO, Ó.; SERVA PAUCAR, D. Recursos cohesivos en textos periodísticos: análisis comparativo en cinco publicaciones del diario El Comercio de Perú. *Lengua y Sociedad*, v. 21, n. 1, p. 403-423, 30 jun. 2022. DOI: <http://dx.doi.org/10.15381/lengsoc.v21i1.22481>.
- PEREIRA, H. B. B.; GRILO, M.; FADIGAS, I. S.; SOUZA JUNIOR, C. T.; CUNHA, M. V.; BARRETO, R. S. F. D.; ANDRADE, J. C.; HENRIQUE, T. Systematic review of the "semantic network" definitions. *Expert Systems with Applications*, v. 1, n. 210, p. 118455, dez. 2022a. DOI: <https://doi.org/10.1016/j.eswa.2022.118455>.
- PEREIRA, H. B. B.; ROSÁRIO, R. S.; PEREIRA, E. J. A. L.; MOREIRA, D. M.; FERREIRA, P.; MIRANDA, J. G. V. Network dynamic and stability on European Union. *Physica A: Statistical Mechanics and its Applications*, v. 587, p. 126532, fev. 2022b. DOI: <https://doi.org/10.1016/j.physa.2021.126532>.
- PETTICREW, M.; ROBERTS, H. *Systematic Reviews in the Social Sciences*. Malden, MA: Blackwell Publishing, 2006.

- PURSER, E.; DREYFUS, S.; JONES, P. Big ideas & sharp focus: researching and developing students' academic writing across the disciplines. *Journal of English for Academic Purposes*, v. 43, p. 100807, jan. 2020. DOI: <https://doi.org/10.1016/j.jeap.2019.100807>.
- ROMAN, A. F.; ROMAN, R. M. The relation discourse-text and textuality. Pro-pragmatic self-reference on speech. *Procedia - Social and Behavioral Sciences*, v. 163, p. 214-219, dez. 2014. DOI: <https://doi.org/10.1016/j.sbspro.2014.12.309>.
- ROSÁRIO, R. S.; CARDOSO, P. T.; MUÑOZ, M. A.; MONTOYA, P.; MIRANDA, J. G. V. Motif-synchronization: a new method for analysis of dynamic brain networks with EEG. *Physica A: Statistical Mechanics and its Applications*, v. 439, p. 7-19, 2015. DOI: <https://doi.org/10.1016/j.physa.2015.07.018>.
- ROTTENSTEINER, S. Structure, function and readability of new textbooks in relation to comprehension. *Procedia - Social and Behavioral Sciences*, v. 2, n. 2, p. 3892-3898, 2010. DOI: <https://doi.org/10.1016/j.sbspro.2010.03.611>.
- SÁNDOR, A. Poeticity. *Poetics*, v. 18, n. 3, p. 299-316, jun. 1989. DOI: [https://doi.org/10.1016/0304-422X\(89\)90005-3](https://doi.org/10.1016/0304-422X(89)90005-3).
- SAUSSURE, F. *Curso de linguística geral*. São Paulo: Cultrix, 2006.
- SEGEV, E. Textual network analysis: detecting prevailing themes and biases in international news and social media. *Sociology Compass*, v. 14, n. 4, p. e12779, abr. 2020. DOI: <https://doi.org/10.1111/soc4.12779>.
- SILVA, B. B. M.; MIRANDA, J. G. V.; CORSO, G.; COPELLI, M.; VASCONCELOS, N.; RIBEIRO, S.; ANDRADE, R. F. S. Statistical characterization of an ensemble of functional neural networks. *The European Physical Journal B*, v. 85, n. 10, p. 358, out. 2012. DOI: <https://doi.org/10.1140/epjb/e2012-30481-7>.
- STANISZ, T.; DROŽDŽ, S.; KWAPIEN, J. Complex systems approach to natural language. *Physics Reports*, v. 1053, p. 1-84, fev. 2024. DOI: <https://doi.org/10.1016/j.physrep.2023.12.002>.
- TEIXEIRA, G. M.; AGUIAR, M. S. F.; CARVALHO, C. F.; DANTAS, D. R.; CUNHA, M. do V.; MORAIS, J. H. M.; PEREIRA, H. B. B.; MIRANDA, J. G. V. Complex semantic networks. *International Journal of Modern Physics C*, v. 21, n. 3, p. 333-347, 2010. DOI: <https://doi.org/10.1142/S0129183110015142>.
- TITSCHER, S.; MEYER, M.; WODAK, R.; VETTER, E. What is a text? In: TITSCHER, S.; MEYER, M.; WODAK, R.; VETTER, E. (Eds.). *Methods of Text and Discourse Analysis*. London: Sage Publications, 2000. p. 20-30.
- WILSON, A. What is a text? *Studies in History and Philosophy of Science Part A*, v. 43, n. 2, p. 341-358, jun. 2012. DOI: <https://doi.org/10.1016/j.shpsa.2011.12.027>.
- YANG, W.; SUN, Y. The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. *Linguistics and Education*, v. 23, n. 1, p. 31-48, mar. 2012. DOI: <https://doi.org/10.1016/j.linged.2011.09.004>.



Este texto está licenciado com uma Licença Creative Commons Atribuição 4.0 Internacional.

APÊNDICE A - CARACTERÍSTICAS CONSIDERADAS NAS DEFINIÇÕES E CARACTERIZAÇÕES

[illegible]

(Huang; Mohan, 2009)	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(Kaplan; Grabe, 2002)	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0
(Khalil, 1989)	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0
(Kitis; Milapides, 1997)	1	0	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
(Kovalchuk; Popivniak, 2021)	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(Luke, 1989)	1	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
(Lyons, 2021)	1	0	1	1	1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0	0
(Makino, 1982)	0	0	0	1	1	1	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0
(Marashi, 2020)	1	0	0	1	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
(Mee; Seng, 2014)	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(Ocak; Baysal, 2016)	1	0	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0
(Okurowski, 1989)	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	1	0	0
(Oliveira; Senna; Pereira, 2024)	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0
(Ouden; Noordman; Terken, 2009)	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
(Pavel, 1973)	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
(Perrin; Petry, 1999)	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
(Perrin; Petry, 2003)	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
(Purser; Dreyfus; Jones, 2020)	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
(Roman; Roman, 2014)	1	1	0	1	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0
(Rottensteiner, 2010)	1	0	0	0	1	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0
(Sándor, 1989)	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
(Stanisz; Drożdż; Kwapień, 2024)	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
(Yang; Sun, 2012)	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0

Notas. 1. Características: (1) coesão; (2) nível lógico ou semântico; (3) função comunicativa; (4) situacionalidade; (5) aceitabilidade; (6) modalidade escrita; (7) modalidade oral; (8) intencionalidade; (9) unidade formal ou estrutural; (10) coerência; (11) nível discursivo; (12) nível lexical; (13) irrelevância do tamanho; (14) caráter linear ou sequencial; (15) informatividade; (16) intertextualidade; (17) nível sintático; (18) distinção de sentenças aleatórias; (19) nível fonológico ou prosódico; (20) estilo; (21) padrões de repetição; (22) possibilidade de ser não-verbal. 2. As definições e caracterizações analisadas podem ser conferidas em Oliveira (2024).

Fonte: Elaborado pelos autores.

APÊNDICE B - AGRUPAMENTO DOS ESTUDOS DE ACORDO COM AS CARACTERÍSTICAS CONSIDERADAS

Grupo	Trabalho	Características do texto ¹																					
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	(Dilidüzgün, 2013)	1	1	1	1	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0
	(Lyons, 2021)	1	0	1	1	1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0	0	0
	(Campuzano; Paucar, 2022)	1	0	1	1	1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0	0	0
	(Brooks, 1997)	0	1	1	0	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1	0	0
	(Dalalyan, 2015)	1	1	1	1	1	0	1	1	0	1	0	1	0	0	1	1	1	0	1	1	0	0
	(Cornish, 2009)	1	1	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	1
	(Roman; Roman, 2014)	1	1	0	1	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0
	(Yang; Sun, 2012)	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
	(Rottensteiner, 2010)	1	0	0	0	1	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0
	(Luke, 1989)	1	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	(Kitis; Milapides, 1997)	1	0	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	(Collet, 2004)	1	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	(Geist, 1987)	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0
	(Fowler, 1979)	0	0	1	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
2	(Khalil, 1989)	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0
	(Okurowski, 1989)	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	1	0	1	0
	(Oliveira; Senna; Pereira, 2024)	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0
	(Purser; Dreyfus; Jones, 2020)	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	(Fang, 1997)	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	(Golden; Pappas, 1990)	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	(Huang; Mohan, 2009)	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	(Albarrán-Santiago, 2015)	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	(Blom; Hansen, 2015)	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	(Cheung, 2010)	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0

	(Kovalchuk; Popivniak, 2021)	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	(Amaral, 1985)	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	(Kaplan; Grabe, 2002)	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1
	(Ouden; Noordman; Terken, 2009)	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	(Pavel, 1973)	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
	(Perrin; Petry, 2003)	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
	(Perrin; Petry, 1999)	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	(Stanisz; Drożdż; Kwapień, 2024)	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
	(Bache, 2010)	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	(Beaugrande, 1980)	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
	(Beaugrande, 1979)	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	(Sándor, 1989)	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
3	(Bayat; Yurdakul, 2014)	0	1	1	0	0	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0
	(Ali, 2021)	1	1	1	0	0	1	1	0	0	1	0	0	1	0	1	0	0	0	0	0	0
	(Felisbino, 2001)	1	1	1	0	1	1	1	0	1	1	1	0	1	0	0	0	0	1	0	0	0
	(Ocak; Baysal, 2016)	1	0	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0
	(Carrell, 1982)	1	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	(Alexander, 2019)	0	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	(Makino, 1982)	0	0	0	1	1	1	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0
	(Marashi, 2020)	1	0	0	1	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
	(Beaugrande, 1983)	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	(Gowhary <i>et al.</i> , 2015)	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	(Haberland, 1999)	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	(Mee; Seng, 2014)	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Notas. 1 Características: (1) coesão; (2) nível lógico ou semântico; (3) função comunicativa; (4) situacionalidade; (5) aceitabilidade; (6) modalidade escrita; (7) modalidade oral; (8) intencionalidade; (9) unidade formal ou estrutural; (10) coerência; (11) nível discursivo; (12) nível lexical; (13) irrelevância do tamanho; (14) caráter linear ou sequencial; (15) informatividade; (16) intertextualidade; (17) nível sintático; (18) distinção de sentenças aleatórias; (19) nível fonológico ou prosódico; (20) estilo; (21) padrões de repetição; (22) possibilidade de ser não-verbal. 2.As definições e caracterizações analisadas podem ser conferidas em Oliveira (2024).

Fonte: Elaborado pelos autores.