REVISTA RAÍDO

OPEN ACCESS

UFGD

DOI: 10.30612/raido.v19i48.20339

Modalidade, polaridade e legibilidade em resumos académicos: uma análise comparativa multidisciplinar

Modality, polarity, and legibility in academic abstracts: a multidisciplinary comparative analysis

Sílvia Araújo[1]
E-mail: saraujo@elach.uminho.pt
Orcid: https://orcid.org/0000-0003-4321-4511

Micaela Aguiar[2]
E-mail: maguiar@letras.up.pt
Orcid: https://orcid.org/0000-0002-5923-9257

Resumo: Este estudo investiga a relação entre modalidade, polaridade e legibilidade em resumos académicos de 40 teses de doutoramento em quatro áreas do conhecimento – Ciências Exatas, Ciências Naturais, Ciências da Saúde e Ciências Sociais e Humanas. Enquadrada numa perspetiva interdisciplinar da Análise do Discurso, em articulação com os contributos do Processamento de Linguagem Natural, a metodologia assenta nestas técnicas para examinar de que forma a legibilidade, a construção discursiva de polaridade (positiva/neutra/negativa) e assim como as modalidades implicadas variam, refletindo as especificidades epistemológicas e metodológicas de cada domínio científico. A legibilidade é avaliada através de índices que medem a facilidade de leitura. A análise do nível de polaridade dos resumos foca-se na aná-

1 Professora na Universidade do Minho, ELACH, CEHUM.
2 Professora na Universidade do Porto, FLUP, CLUP.

lise lexical de valores positivos, neutros ou negativos. A dimensão da modalidade centra-se na identificação de elementos modais que expressam as diferentes modalidades (epistémica, deôntica e desiderativa) presentes nos resumos. Ao analisar estas três dimensões, o estudo oferece uma perspetiva holística da escrita académica, permitindo mapear padrões discursivos que caracterizam as diferentes áreas do saber.

Palavras-chave: legibilidade; positividade; modalidade; resumo de tese de doutoramento, discurso académico

Abstract: This study investigates the relationship between modality, polarity, and readability in academic abstracts from 40 doctoral theses across four areas of knowledge – Exact Sciences, Natural Sciences, Health Sciences, and Social and Human Sciences. The methodology is based on advanced natural language processing techniques to examine how readability, the discursive construction of polarity (positive/neutral/negative), and the involved modalities vary, reflecting the epistemological and methodological specificities of each scientific domain. Readability is assessed through indices that measure ease of reading. The analysis of the level of polarity in the abstracts focuses on the lexical analysis of positive, neutral, or negative values. The modality dimension centers on identifying modal elements that express the different modalities (epistemic, deontic, and desiderative) present in the abstracts. By analyzing these three dimensions, the study offers a holistic perspective on academic writing, allowing for the mapping of discursive patterns that characterize different areas of knowledge.

Keywords: readability; positivity; modality; PhD thesis abstract; academic discourse

1 INTRODUÇÃO

O presente estudo exploratório tem como objetivo analisar a modalidade, a polaridade e a legibilidade em resumos de teses de doutoramento em quatro áreas disciplinares, as Ciências Exatas, as Ciências Naturais, as Ciências da Saúde e as Ciências Sociais e Humanas. Estas três dimensões aparentemente distintas, intersecccionam-se nas expectativas de produção dos géneros académicos (Coutinho & Miranda, 2009), como a expetativa de que os géneros académicos sejam “objetivos”, que não tenham marcas dos posicionamentos do locutor ou demonstrem indícios de uma orientação mais ou menos positiva, assim como a expetativa de que sejam necessariamente difíceis de ler, distanciando a comunidade académica da não-académica. Será que estas expectativas se materializam efetivamente nas produções reais deste género académico? Que distinções entre áreas disciplinares existirão em termos de cada uma das três dimensões. São estas as questões de investigação que motivam e orientam o nosso estudo.

2 ENQUADRAMENTO TEÓRICO

O presente trabalho insere-se no quadro teórico de uma perspetiva interdisciplinar da Análise do Discurso, em articulação com os contributos do Processamento de Linguagem Natural. Neste enquadramento, descrevemos os conceitos centrais deste estudo: o género académico “resumo de tese de doutoramento” e os conceitos de “modalidade”, “polaridade” e “legibilidade”, especialmente no que ao discurso académico diz respeito.

2.1 O género académico “resumo de tese de doutoramento”

O género académico “resumo de tese de doutoramento” funciona como uma porta de entrada para a investigação integral. De acordo com Silva e Chaves (2017), este género tem, entre outras funções, a de persuadir o leitor a prosseguir a leitura da tese, funcionando como um convite à exploração mais aprofundada do trabalho. Do ponto de vista da organização textual, Santos e Silva (2018) consideram-no um género incluído, que surge também nas dissertações de mestrado e nos artigos científicos,

e um género peritextual, localizado sequencialmente antes do corpo da tese. Do ponto de vista da estrutura, modelos propostos por investigadores como Bhatia (1993) e Hyland (2002) oferecem enquadramentos úteis para a análise de resumos académicos. Segundo Carvalho (2010), Bhatia (1993), por exemplo, propõe que os resumos de artigos científicos devem incluir quatro elementos essenciais: o que o autor fez, como o fez, o que encontrou e o que concluiu. Por sua vez, Hyland (2002) sugere um modelo constituído por cinco movimentos retóricos: (1) introdução, (2) propósito, (3) método, (4) produto (ou resultados) e (5) conclusão. Embora ambos os modelos tenham sido concebidos originalmente para resumos de artigos académicos, a sua aplicabilidade ao género resumo de tese tem sido considerada em diversos estudos, dada a proximidade funcional e estrutural entre os dois tipos de texto. Diversos estudos têm incidido sobre partes distintas das teses e dissertações, como as introduções ou as conclusões; contudo, como sublinha El-Dakhs (2018), são menos numerosos os trabalhos que se debruçam especificamente sobre os resumos de teses de doutoramento. Também em Portugal, a análise deste género específico ainda se encontra pouco explorada, como assinalam Santos e Silva (2016), o que justifica a relevância da escolha deste género como objeto de estudo do presente trabalho.

2.2 Modalidade, Polaridade e Legibilidade no discurso académico

A modalidade é um conceito de uma grande heterogeneidade de abordagens teóricas (Marques, 2014). As principais teorizações em torno da modalidade, como a de Campos (2004), assentam em três valores modais: o valor epistémico, ligado a um posicionamento relativo a contínuos de certeza a impossibilidade, o valor deôntico, correspondente a dimensões de desejabilidade e de obrigatoriedade e o valor apreciativo ou desiderativo, relativo a juízos avaliativos em relação a um determinado estado de coisas. Nascimento (2018), numa análise da modalidade em diversos géneros académicos, observa que a modalidade epistémica (com diferentes valores) e a modalidade apreciativa surgem em todos os géneros estudados, como o artigo científico, o projeto de dissertação, a monografia, incluindo o resumo, contudo a modalidade deôntica surge apenas em resumos e protocolos académicos administrativos. Também Silva e Chaves (2017), num estudo sobre o resumo académico de artigos científicos e atas, observam a predominância dos valores epistémico e apreciativos neste género. Como afirmam Silva e Chaves (2017, p. 11), a inscrição da subjetividade do locutor através de estratégias de modalidade “amplia a discussão sobre a neutralidade, a impessoalidade e a imparcialidade dos textos considerados ‘objetivos’”.

No mesmo sentido, considerar os géneros académicos num eixo de polaridade mais ou menos negativo ou positivo pode contribuir, não só para revelar padrões diferenciais entre áreas disciplinares, mas apontar também para a influência de condicionantes contextuais na orientação argumentativa global deste género. A polaridade é uma métrica de Processamento de Linguagem Natural, que categoriza unidades discursivas, maiores ou menores, em termos do sentimento positivo, neutro ou negativo e que pode ser medida através de diversas abordagens, cada uma com as suas limitações inerentes. Numa abordagem com base em “machine learning”, a análise de sentimento requer uma grande quantidade de dados para obter resultados bons (Liu & Zhu,. 2023); já numa abordagem assente no léxico, os valores de sentimento de um texto são calculados através de base lexical composta por palavras que já estão etiquetadas com pontuações de sentimento (Zhang, 2014). Wen & Lei (2022) verificaram que a abordagem assente no léxico apresenta resultados satisfatórios na avaliação da polaridade de géneros académicos. Atualmente existem também abordagens baseadas em LLMs. Por exemplo, Wang et al. (2023) focaram-se nas capacidades do ChatGPT em termos de análise de sentimento em comparação a modelos treinados com data categorizada, como os modelos BERT e Sota. Os autores verificaram que o ChatGPT produziu resultados equiparáveis a esses modelos, mesmo em contexto de “zero shot” (isto é, sem exemplos ilustrativos integrados no *prompt*).

A fórmula de Flesch, conhecida como Índice de Leitura de Flesch, é um dos instrumentos de avaliação de legibilidade mais utilizados. Baseia-se no cálculo do comprimento médio das frases e do número médio de sílabas por palavra, e tem como objetivo ajudar professores e educadores a avaliarem corretamente a complexidade dos materiais didáticos antes da sua aplicação em contexto de sala de aula (Imperial & Madabushi, 2023). Outras medidas têm sido propostas para colmatar as limitações da fórmula Flesch, incluíndo o facto de ter sido concebida para o Inglês, como, por exemplo, o Índice Gunning Fog, medido com base no comprimento das frases e no uso de palavras longas de três ou mais sílabas, o Índice de Legibilidade Automatizado (Automated Readability Index), medido com base no número de caracteres por palavra e palavras por frase, e o Índice de Coleman-Liau (Coleman-Liau Index), medido com base na quantidade de letras por 100 palavras e número de frases por 100 palavras (Silveira et al., 2024). Os textos académicos apresentam, sem surpresas, níveis elevados de complexidade, tornando-os particularmente difíceis de ler (Rodrigo et al., 2024).

3 METODOLOGIA

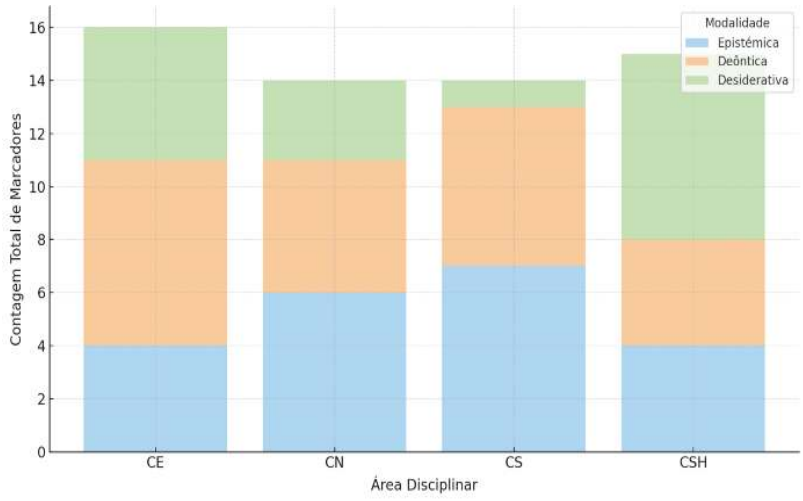
Para a constituição do *corpus* que serviu de objeto de estudo, recolhemos 40 teses de doutoramento de repositórios portugueses de universidades públicas (como o RepositóriUM da Universidade do Minho e o Repositório Aberto da Universidade do Porto) 10 de cada uma das seguintes áreas disciplinares, Ciências Exatas, Ciências Naturais, Ciências da Saúde e Ciências Sociais e Humanas. Como critérios de seleção, determinámos que as teses deveriam encontrar-se em acesso aberto, estarem escritas em Português Europeu e terem sido produzidas nos últimos 20 anos. Após a recolha do *corpus*, extraímos apenas os resumos. Para a análise da modalidade, faremos uma análise lexical, que parte da lista de verbos, advérbios e adjetivos modais tal como se encontram descritos na Gramática do Português Vol. 1 (Raposo et al., 2013, p. 623-668), considerando os contributos de estudos sobre modalidade no discurso académico, como Nascimento & Geziel de Brito (2012). São eles: (1) Verbos modais: dever, poder, (2) Verbos plenos: obrigar, permitir, autorizar, desejar, pretender, procurar, querer, esperar, saber, (3) Advérbios: possivelmente, provavelmente, eventualmente, necessariamente, certamente, dificilmente, sem dúvida, duvidosamente, forçosamente, indiscutivelmente, desejavelmente, obrigatoriamente e (4) Adjetivos: é necessário, é desejável, é provável, é essencial, é crucial, é importante, é possível. A extração dos verbos, advérbios e adjetivos modais foram automatizados num *script* em Python, que recorreu à biblioteca pandas para a gestão de dados estruturados e a ferramentas adicionais de processamento linguístico. A polaridade foi analisada com recurso à API da OpenAI para classificar a polaridade dos resumos, permitindo assim uma avaliação qualitativa automatizada da orientação avaliativa dos resumos. A análise da legibilidade foi realizada através do software ALT (Análise de Legibilidade Textual), que calcula métricas de legibilidade com base nos índices de Flesch-Kincaid (FK), Gunning Fog (GF), Automated Readability Index (ARI) e Coleman-Liau (CL), e que foi desenvolvido a partir de métricas de legibilidade originais adaptadas para a Língua Portuguesa (Moreno et al. 2022). A ferramenta conta o número de letras, de sílabas, de palavras, de frases e de palavras complexas e fornece o resultado final da legibilidade de um texto através da média aritmética destes quatro índices.

4 RESULTADOS E DISCUSSÃO

4.1 Modalidade

A Figura 1 apresenta a distribuição total de ocorrências dos três valores modais considerados, nomeadamente o valor epistémico, deôntico e desiderativo, nos resumos analisados, organizada por área disciplinar.

Figura 1. Distribuição dos valores modais por área disciplinar



Fonte: Figura produzida pelas autoras.

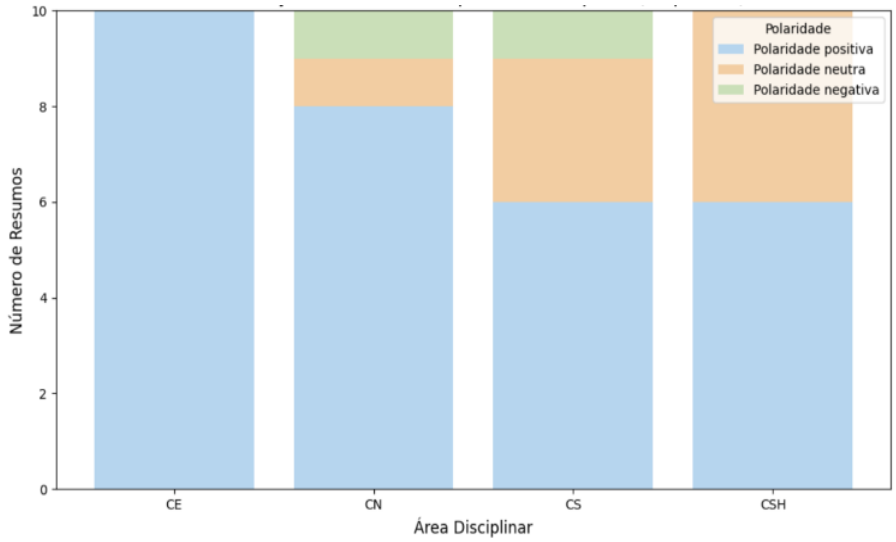
Nas Ciências Exatas (CE) e nas Ciências Naturais (CN), há predominância das modalidades deôntica e epistémica. Estes resultados não refletem conclusões de estudos anteriores para resumos de artigos científicos. Vázquez e Giner (2008) concluíram no seu trabalho comparativo entre resumos de artigos de Marketing, Biologia e Engenharia Mecânica, que Marketing apresentava o maior número de ocorrências da modalidade epistémica, seguido de Biologia com metade das ocorrências e de Engenharia Mecânica com ainda menos ocorrências. Tal pode sugerir uma distinção em termos de valores modais entre resumos de teses de doutoramento e resumos de artigos científicos. Nas Ciências da Saúde (CS), a modalidade epistémica é a predominante. Montkhongtham (2021), que examinou resumos de artigos científicos na área da Medicina, aponta para a mesma conclusão, sublinhado o facto de que

se trata de uma disciplina que lida com muita incerteza. Já nas Ciências Sociais e Humanas (CSH), observa-se um padrão distinto: a modalidade desiderativa é a mais frequente. A predominância da modalidade desiderativa na área das Ciências Sociais e Humanas encontra-se em parte descrita no trabalho de Nascimento e Geziel de Brito (2012) centrado em resumos de eventos científicos na área da Linguística e do Ensino de Línguas. Os autores indicam que a modalidade desiderativa é a segunda com mais ocorrências, precedida pela modalidade epistémica, como a modalidade com mais ocorrências. Estes dados revelam que cada área mobiliza as modalidades de forma própria, em consonância com suas práticas discursivas e epistemológicas. Por exemplo, em CE e CN, a combinação de modalidade deôntica e epistémica poderá ser 'induzida' por uma estrutura discursiva fortemente normatizada no que diz respeito à metodologia (ex.: descrição de procedimentos, critérios e limitações), que pedem marcadores de obrigação/procedimento ('deve', 'é necessário') para sinalizar replicabilidade, e de incerteza controlada ('pode', 'sugere-se', 'é provável') para calibrar a força das inferências. Já em CS, a predominância epistémica poderá estar alinhada com a gestão intrínseca da incerteza clínica e estatística: há pressão ética e editorial para evitar sobre-afirmações e para explicitar risco, variabilidade e evidência, o que favorece atenuadores e qualificação de resultados. Já em CSH, a maior frequência da modalidade desiderativa poderá ser resultante de práticas de investigação orientadas para a problematização e para a intervenção social/cultural; nesses resumos é recorrente projetar objetivos, alcances e implicações ('pretende-se', 'almeja-se', 'busca-se'), convidando o leitor a aderir a um programa interpretativo. Acresce que resumos de teses tendem a enfatizar percurso e requisitos (o que favorece marcas deônticas), enquanto resumos de artigos priorizam resultados e cautela interpretativa (o que favorece marcas epistémicas), o que pode contribuir para a distinção modal entre géneros e áreas.

4.2 Polaridade

A Figura 2 mostra os resultados da distribuição da polaridade por área disciplinar. Os resultados revelam que os resumos de teses de doutoramento apresentam uma polaridade maioritariamente positiva, especialmente nas áreas das Ciências Exatas e das Ciências Naturais.

Figura 2. Distribuição da polaridade por área disciplinar



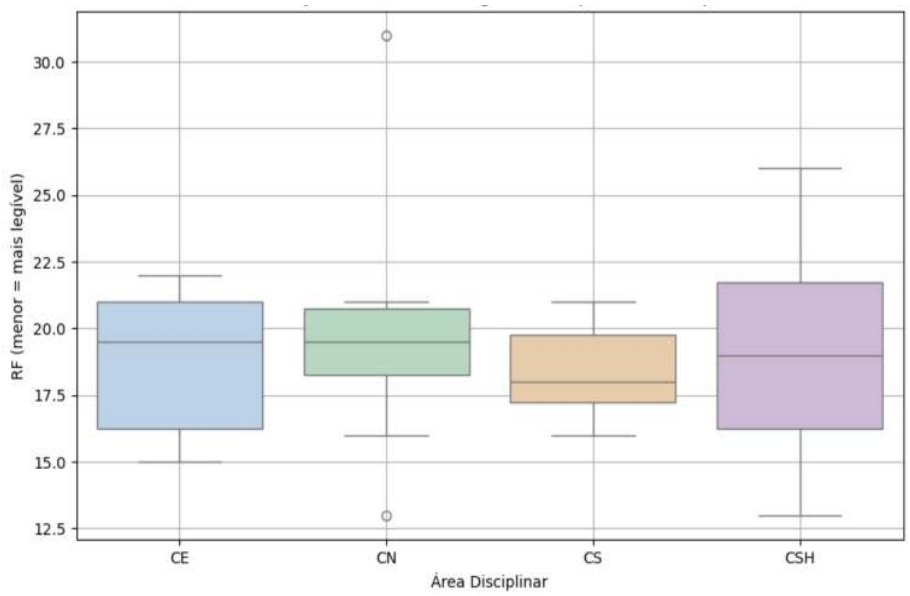
Fonte: Figura produzida pelas autoras.

Estes resultados confirmam as conclusões de Liu & Zhu (2023), que analisaram a positividade linguística em oito disciplinas (Comunicação, Linguística, Ciência Política, Sociologia, Indústria aeroespacial, Automação, Engenharia de software e Transportes). Os autores concluíram não só que as ciências duras apresentam um grau mais elevado de positividade linguística, mas que, ao longo do tempo, as ciências duras registaram taxas crescentes de positividade linguística. Vinkers et al. (2015), que investigaram o recurso a palavras positivas e negativas nos títulos e resumos de artigos médicos publicados num espaço de 40 anos, também revelaram que a frequência de palavras positivas aumentou mais rapidamente do que a de palavras negativas nos resumos médicos; não estando este aumento relacionado com tendências linguísticas gerais, nem com o uso de algumas palavras específicas, uma vez que todas as palavras positivas registaram um aumento de frequência. Também Wen & Lei (2022) identificam um viés positivo na escrita académica, com a análise de um corpus de 775 460 resumos publicados entre 1969 e 2019 em 123 revistas científicas que abrangem 12 disciplinas de investigação em ciências da vida. Entre as razões apontadas para explicar este fenómeno são mencionadas a preferência de revistas científicas por resultados positivos e igualmente o emprego de estratégias discursivas para a promoção dos resultados de investigação por parte dos autores.

4.3 Legibilidade

Recorrendo ao software ATL, os resumos foram classificados em três categorias: (1) Alta legibilidade: RF < 13, (2) Média legibilidade: 13 ≤ RF < 17 e (3) Baixa legibilidade: RF ≥ 17. A Figura 3 apresenta a distribuição dos valores do índice RF (Resultado Final de legibilidade) por área disciplinar.

Figura 3. Distribuição da métrica de legibilidade RF por área disciplinar



Fonte: Figura produzida pelas autoras.

A área das Ciências Exatas apresenta uma mediana próxima de 19, indicando que a maioria dos resumos desta área possui legibilidade média a baixa. Há uma concentração de resumos entre 15 e 22, sugerindo variabilidade no grau de legibilidade dos resumos, mas sem valores extremos significativos. Também a área das Ciências Naturais exibe uma distribuição similar, com valores centrados em torno de 20, indicando também baixa legibilidade para a maioria dos resumos. A presença de um *outlier* (próximo a 31) sugere um texto com legibilidade consideravelmente baixa. A área das Ciências da Saúde mostra valores ligeiramente mais baixos, com a mediana em torno de 18, sugerindo que há uma tendência para textos mais legíveis comparados às áreas anteriores. Já a área das Ciências Sociais e Humanas apresenta uma maior dispersão de dados, variando de 13 a 25. A mediana é próxima de 18, mas o

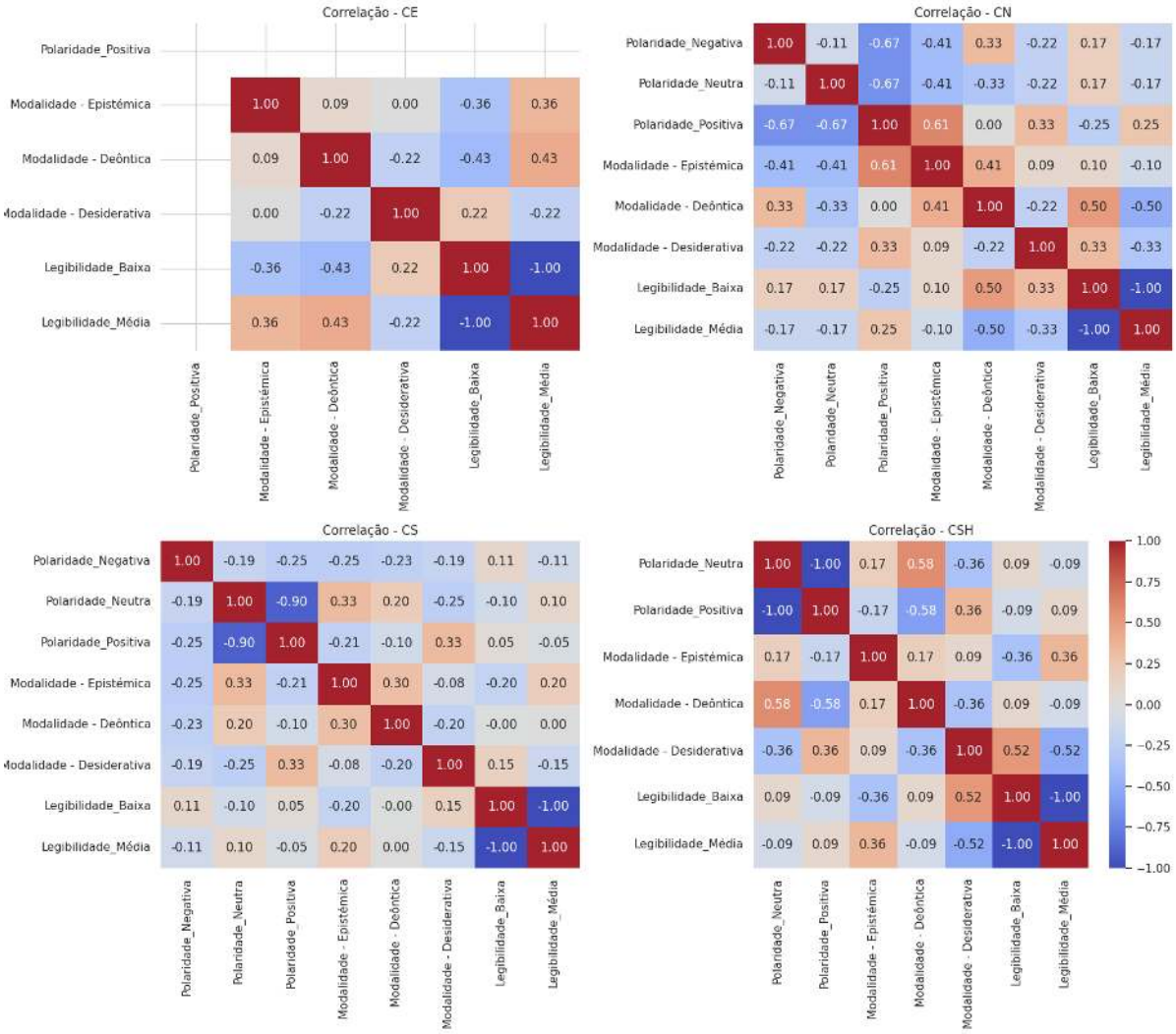
intervalo interquartil revela que uma parcela significativa dos resumos possui legibilidade média a baixa, com alguns resumos com níveis maiores de legibilidade (valores abaixo de 13).

Os resultados vão ao encontro da literatura sobre legibilidade em géneros académicos. Gazni (2011), que analisou resumos de artigos científicos das cinco instituições mais citadas do mundo, conclui que a medicina, a química e domínios relacionados, têm os resumos mais difíceis de ler. Também Ante (2022) revela que, no domínio das tecnologias, os resumos de artigos científicos têm-se tornado mais difíceis de ler. Aliás, ambos os autores referem que artigos com resumos com índices de menor legibilidade são mais citados e Metoyer-Duran (1993) constatava já nos anos 90 que os artigos rejeitados eram mais legíveis do que os aceites. Também nas Ciências Sociais e Humanas os resumos têm vindo a tornar-se menos legíveis, como comprovam Wang et al. (2022) para a área das línguas e da linguística. Apesar da valorização crescente de práticas de divulgação científica por entidades de financiamento científico, esta tendência, ainda que incidente sobretudo em artigos científicos e os seus resumos, cria o “incentivo perverso”, dadas as pressões que os investigadores sofrem para a publicação constante (cristalizadas na conhecida expressão “publish or perish”), para que estes tornem os seus textos mais difíceis de ler e consequentemente menos acessíveis ao público em geral.

4.4 Correlações entre modalidade, polaridade e legibilidade por área disciplinar

Numa perspetiva exploratória, decidimos analisar as correlações possíveis entre as dimensões em análise a modalidade (epistémica, deôntica e desiderativa), a polaridade (positiva, neutra e negativa) e a legibilidade (alta, média e baixa) nas quatro áreas disciplinares estudadas. Na Figura 4, encontra-se representado um mapa de calor das correlações possíveis. Um mapa de calor de correlações mostra graficamente a força e a direção das relações lineares entre variáveis, usando uma escala de cores para representar os coeficientes de correlação (r), que variam de -1 a +1. Valores próximos de +1 indicam uma correlação positiva forte, valores próximos de -1 indicam uma correlação negativa forte e valores próximos de 0 indicam ausência de correlação linear significativa. Cores mais intensas correspondem a correlações mais fortes, enquanto tons neutros indicam relações fracas ou inexistentes. A diagonal principal, com r = 1, reflete a autocorrelação de cada variável.

Figura 4. Correlações entre modalidade, polaridade e legibilidade por área disciplinar



Fonte: Figura produzida pelas autoras.

A análise das correlações positivas iguais ou superiores a 0.30 (≥ 0.30) entre as variáveis permitiu identificar padrões linguístico-discursivos distintos nas quatro áreas disciplinares estudadas. A abordagem baseou-se exclusivamente nos dados extraídos dos heatmaps, preservando a terminologia empregue nas visualizações originais.

Nas Ciências Exatas, observaram-se associações relevantes entre legibilidade média e duas modalidades discursivas: a modalidade deôntica (0.43) e a modalidade epistémica (0.36). A dupla correspondência entre legibilidade média e modalidade deôntica (em ambos os sentidos) sugere que, nesta área, os resumos que formulam recomendações ou prescrições de procedimentos tendem a apresentar uma maior legibilidade, ainda que tecnicamente densa. A correlação da modalidade epistémica e da legibilidade aponta para uma valorização de interpretações possíveis e não definitivas dos resultados que contribuem para a acessibilidade do resumo.

Nas Ciências Naturais, os dados evidenciam uma forte relação entre polaridade positiva e modalidade epistémica (0.61), o que pode refletir uma tendência para construir discursos sobre o conhecimento científico de forma positiva, embora com modalização dessa positividade através de valores não compromissivos de certeza. Além disso, verifica-se uma correlação positiva entre a modalidade deôntica e legibilidade baixa (0.50), sugerindo que, nesta área, os enunciados que expressam necessidade de ação, de adoção de procedimentos ou de seguimento de diretrizes assumem formas mais complexas ou tecnicamente densas. A modalidade epistémica também se associa à deôntica (0.41), o que indica uma articulação frequente entre interpretações baseadas em dados e propostas de atuação ou diretrizes.

Nas Ciências da Saúde, destaca-se a associação entre polaridade positiva e modalidade desiderativa (0.33), o que sugere que os enunciados em que os autores expressam intenções investigativas ou projeções futuras tendem a ser apresentados de forma positiva. A correlação entre modalidade deôntica e modalidade epistémica (0.30) reforça a ideia de que, neste domínio, é comum que se articulem interpretações baseadas em dados empíricos com afirmações sobre necessidades práticas, recomendações clínicas ou exigências de intervenção, reforçando a dimensão aplicada do discurso científico na saúde.

Nas Ciências Sociais e Humanas, destacam-se quatro correlações. A mais forte ocorre entre polaridade neutra e modalidade deôntica (0.58), sugerindo uma elevada frequência de enunciados que exprimem necessidade ou prescrição teórica sem recorrer a uma marcação de polaridade explícita. A associação entre legibilidade baixa e modalidade desiderativa (0.52) pode indicar que as expressões de finalidade dos estudos ocorrem frequentemente em enunciados com maior complexidade linguística, provavelmente marcados por referências teóricas. A modalidade desiderativa também apresenta relação com polaridade positiva (0.36), o que sugere que os enunciados

nos quais se manifestam objetivos tendem a ser expressos de forma positiva. Finalmente, a modalidade epistémica mostra-se associada à legibilidade média (0.36).

5 CONCLUSÕES

O presente estudo permitiu identificar padrões distintos de modalidade, polaridade e legibilidade nos resumos de teses de doutoramento em quatro grandes áreas disciplinares: Ciências Exatas, as Ciências Naturais, as Ciências da Saúde e as Ciências Sociais e Humanas.Verificou-se que, nas Ciências Exatas e Naturais, prevalecem as modalidades deôntica e epistémica, enquanto nas Ciências da Saúde predomina a modalidade epistémica, e nas Ciências Sociais e Humanas destaca-se a modalidade desiderativa. Estes resultados sugerem que as teses, enquanto género académico específico, mobilizam valores modais de forma diferenciada em relação aos artigos científicos, o que pode refletir práticas discursivas e epistemológicas próprias de cada área. No que diz respeito à polaridade, os resultados indicam uma tendência para o recurso a linguagem positiva, sobretudo nas áreas mais técnicas e experimentais, corroborando estudos anteriores que apontam para um viés positivo crescente na escrita académica. Esta tendência pode estar relacionada com estratégias discursivas de promoção dos resultados e com as preferências editoriais por resultados positivos. Quanto à legibilidade, os resumos analisados apresentaram, de forma geral, índices médios a baixos, com maior dificuldade de leitura nas Ciências Exatas e Ciências Naturais. Este resultado confirma a tendência observada na literatura de que os textos académicos, incluindo resumos, tendem a ser pouco acessíveis ao público não especializado, apesar do crescente apelo à divulgação científica. As correlações identificadas entre legibilidade, polaridade e modalidade revelam que estas dimensões não operam de forma isolada, mas antes se articulam para construir estilos discursivos característicos de cada área disciplinar. Observa-se que a modalidade parece influenciar tanto a legibilidade como a polaridade. As modalidades deôntica e desiderativa, associadas a recomendações e objetivos, tendem a relacionar-se com níveis mais baixos de legibilidade, o que sugere que são frequentemente associados a uma maior complexidade linguística. Por outro lado, a modalidade epistémica, ligada a interpretações, hipóteses e avaliações de certeza, surge frequentemente associada a níveis médios de legibilidade, sugerindo que a cautela na formulação de afirmações pode

favorecer uma maior clareza e acessibilidade do texto. Além disso, a associação entre modalidades e polaridade sugere que a construção discursiva do conhecimento científico é frequentemente acompanhada por estratégias que modulam a certeza e o otimismo, adaptadas às práticas epistemológicas e comunicativas de cada domínio.

Embora o presente estudo tenha contribuído para compreender as relações entre legibilidade, polaridade e modalidade nos resumos de teses de doutoramento, algumas limitações devem ser consideradas, nomeadamente, a abordagem quantitativa limita a interpretação de fenómenos mais subtis que uma abordagem qualitativa poderia revelar. Para investigações futuras, seria pertinente complementar com análises qualitativas, assim como alargar o *corpus*, por exemplo, a diferentes géneros académicos incluídos das teses de doutoramento, como a introdução, a metodologia ou as conclusões.

FINANCIAMENTO

Este trabalho é financiado por fundos nacionais através da FCT – Fundação para a Ciência e a Tecnologia, I.P., no âmbito do projeto UID/00022/2025.

Este trabalho é financiado por fundos nacionais através da FCT – Fundação para a Ciência e a Tecnologia, I.P., no âmbito do projeto UID/00305/2025.

REFERÊNCIAS

ANTE, L. The relationship between readability and scientific impact: Evidence from emerging technology discourses. *Journal of Informetrics*, v. 16, n. 1, p. 101252, 2022. Disponível em: https://doi.org/10.1016/j.joi.2022.101252. Acesso em: 06 jul. 2025.

BHATIA, V. K. *Analysing genre: Language use in professional settings*. New York: Longman Publishing, 1993.

CAMPOS, M.ª Henriqueta Costa. A modalidade apreciativa: uma questão teórica. In: OLIVEIRA, Fátima; DUARTE, Isabel Margarida (org.). *Da língua e do discurso: homenagem a Joaquim Fonseca*. Porto: Campo das Letras, 2004. p. 265-281.

CARVALHO, F. F. Padrões de organização textual e lexicogramatical do gênero acadêmico resumo de tese: um estudo de caso. *Trabalhos em Linguística Aplicada*, v. 49, p. 115–128, 2010. Disponível em: https://doi.org/10.1590/s0103-18132010000100009. Acesso em: 06 jul. 2025.

COUTINHO, M. A.; MIRANDA, F. To describe textual genres: problems and strategies. In: BAZERMAN, C.; BONINO, A.; FIGUEIREDO, D. (org.). *Genre in a changing world*. Fort Collins: The WAC Clearinghouse, 2009. Disponível em: https://doi.org/10.37514/per-b.2009.2324.2.03. Acesso em: 06 jul. 2025.

EL-DAKHS, D. A. S. Why are abstracts in PhD theses and research articles different? A genre-specific perspective. *Journal of English for Academic Purposes*, v. 36, p. 48–60, 2018. Disponível em: https://doi.org/10.1016/j.jeap.2018.09.005. Acesso em: 06 jul. 2025.

GAZNI, A. Are the abstracts of high impact articles more readable? Investigating the evidence from top research institutions in the world. *Journal of Information Science*, v. 37, n. 3, p. 273–281, 2011. Disponível em: https://doi.org/10.1177/0165551511401658. Acesso em: 06 jul. 2025.

HYLAND, K. *Teaching and researching writing*. Harlow, UK: Pearson Education Limited, Longman, 2002. Disponível em: https://doi.org/10.4324/9781003198451-8. Acesso em: 06 jul. 2025.

IMPERIAL, J. M.; MADABUSHI, H. T. Flesch or fumble? Evaluating readability standard alignment of instruction-tuned language models. *arXiv preprint*, arXiv:2309.05454, 2023.

LIU, X.; ZHU, H. Linguistic positivity in soft and hard disciplines: temporal dynamics, disciplinary variation, and the relationship with research impact. *Scientometrics*, v. 128, n. 5, p. 3107-3127, 2023. Disponível em: https://www.researchgate.net/publication/369421953_Linguistic_positivity_in_soft_and_hard_disciplines_temporal_dynamics_disciplinary_variation_and_the_relationship_with_research_impact. Acesso em: 06 jul. 2025.

MARQUES, M. A. Linguagem coloquial e modalização. *Redis: Revista de Estudos do Discurso*, n. 3, p. 94–106, 2014.

METOYER-DURAN, C. The readability of published, accepted, and rejected papers appearing in College & Research Libraries. *College & Research Libraries*, v. 64, n. 6, p. 517–526, 1993. Disponível em: https://doi.org/10.5860/crl_54_06_517. Acesso em: 06 jul. 2025.

MONTKHONGTHAM, N. Medical uncertainty and the art of communication: Exploring modality applied in medical journal abstracts. *LEARN Journal: Language Education and Acquisition Research Network*, v. 14, n. 1, p. 604–646, 2021.

MORENO, G. C. D. L.; SOUZA, M. P.; HEIN, N.; HEIN, A. K. ALT: um software para análise de legibilidade de textos em língua portuguesa. *arXiv preprint*, arXiv:2203.12135, 2022.

NASCIMENTO, E. P.; BRITO, L. I. M. A. Os modalizadores como estratégia argumentativa no gênero resumo acadêmico. *Revista Letr@ Viv@*, v. 11, n. 1, p. 55–64, 2012. Disponível em: https://periodicos.ufpb.br/index.php/lv/article/view/15319. Acesso em: 06 jul. 2025.

NASCIMENTO, E. Discursive modality as argumentative index in academic genres. **Fórum Lingüístico**, v. 15, n. 4, 2018.

RAPOSO, E. B. P.; NASCIMENTO, M. F. B.; MOTA, M. A. C.; SEGURA, L.; MENDES, A. *Gramática do português: volume I*. Lisboa: Fundação Calouste Gulbenkian, 2013.

RODRIGO, T. V.; JONATHAN, V. M.; JHON, A. C. Analyzing readability indices in scholarly and non-scholarly texts: Assessing easability and difficulty. *Revista Ibérica de Sistemas e Tecnologias de Informação*, n. E73, p. 644–660, 2024.

SANTOS, J. V.; SILVA, P. N. Issues of textual hybridity in a major academic genre: PhD dissertations vs. research articles. *Redis: Revista de Estudos do Discurso*, n. 5, p. 171–194, 2016.

SANTOS, J. V.; SILVA, P. N. Polifonia na voz autoral: resumo e agradecimentos na tese de doutoramento. In: VELOSO, J. et al. (org.). *A linguística em diálogo: volume comemorativo dos 40 anos do Centro de Linguística da Universidade do Porto*. Porto: FLUP/CLUP, 2018. p. 395–413.

SILVA, M. D.; CHAVES, A. L. A. Os elementos modalizadores e a construção da argumentação no interior do gênero resumo acadêmico. In: *Simpósio Nacional de Linguagens e Gêneros Textuais*, 4., 2017. Disponível em: https://www.editorarealize.com.br/editora/anais/sinalge/2017/TRABALHO_EV066_MD1_SA16_ID6_13022017192503.pdf. Acesso em: 06 jul. 2025.

SILVEIRA, V. I.; MENEZES, P. H.; SILVA, M. S.; CARMO, F. A.; LOBATO, F. M. Classificação de linguagem simples: uma abordagem baseada em leiturabilidade e legibilidade. In: *Workshop de Computação Aplicada em Governo Eletrônico (WCGE)*. SBC, 2024. p. 99–110. Disponível em: https://doi.org/10.5753/wcge.2024.2536. Acesso em: 06 jul. 2025.

VÁZQUEZ ORTA, I.; GINER, D. Beyond mood and modality: Epistemic modality markers as hedges in research articles. A cross-disciplinary study. *Revista Alicantina de Estudios Ingleses*, n. 21, p. 171–190, 2008. Disponível em: https://doi.org/10.14198/raei.2008.21.10. Acesso em: 06 jul. 2025.

VINKERS, C. H.; TIJDINK, J. K.; OTTE, W. M. Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: retrospective analysis. *BMJ*, v. 351, 2015. Disponível em: https://doi.org/10.1136/bmj.h6467. Acesso em: 06 jul. 2025.

WANG, S.; LIU, X.; ZHOU, J. Readability is decreasing in language and linguistics. *Scientometrics*, v. 127, n. 8, p. 4697–4729, 2022. Disponível em: https://doi.org/10.1007/s11192-022-04427-1. Acesso em: 06 jul. 2025.

WANG, Z.; XIE, Q.; FENG, Y.; DING, Z.; YANG, Z.; XIA, R. Is ChatGPT a good sentiment analyzer? A preliminary study. *arXiv preprint*, arXiv:2304.04339, 2023. Disponível em: https://arxiv.org/abs/2304.04339. Acesso em: 13 out. 2025.

WEN, J. U.; LEI, L. Linguistic positivity bias in academic writing: A large-scale diachronic study in life sciences across 50 years. *Applied Linguistics*, v. 43, n. 2, p. 340–364, 2022. Disponível em: https://doi.org/10.1093/applin/amab037. Acesso em: 06 jul. 2025.

ZHANG, H.; GAN, W.; JIANG, B. Machine learning and lexicon based methods for sentiment classification: A survey. In: *2014 11th Web Information System and Application Conference (WISA)*. IEEE, 2014. p. 262–265. Disponível em: https://doi.org/10.1109/WISA.2014.55. Acesso em: 06 jul. 2025.