

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO**

**PUC-SP**

**NATALIA LURI ARIMORI RIBEIRO**

**PORTAL MULTIMODAL/MULTILÍNGUE PARA O AVANÇO DA CIÊNCIA  
ABERTA NAS HUMANIDADES: LINGUAGEM TELE CINEMÁTICA COMO  
RECURSO DE ENSINO DE INGLÊS COMO LÍNGUA ESTRANGEIRA ATRAVÉS  
DE N-GRAMA DE CLASSE SEMÂNTICA (NGCS)**

**MESTRADO EM LINGUÍSTICA APLICADA E ESTUDOS DA LINGUAGEM**

**SÃO PAULO**

**2023**

**NATALIA LURI ARIMORI RIBEIRO**

**PORTAL MULTIMODAL/MULTILÍNGUE PARA O AVANÇO DA CIÊNCIA  
ABERTA NAS HUMANIDADES: LINGUAGEM TELE CINEMÁTICA COMO  
RECURSO DE ENSINO DE INGLÊS COMO LÍNGUA ESTRANGEIRA ATRAVÉS  
DE N-GRAMA DE CLASSE SEMÂNTICA (NGCS)**

Dissertação apresentada à Banca Examinadora da  
Pontifícia Universidade Católica de São Paulo, como exigência  
parcial para obtenção do título de Mestre em Linguística  
Aplicada e Estudos da Linguagem sob orientação do Professor  
Doutor Antonio Paulo Berber Sardinha.

**SÃO PAULO**

**2023**

Autorizo, exclusivamente para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação de mestrado, por processos fotocopiadores ou eletrônicos, desde que devidamente citada.

Assinatura:

Data:

## **FICHA CATALOGRÁFICA**

RIBEIRO, Natalia Luri Arimori. São Paulo, 2023

Dissertação de Mestrado – Pontifícia Universidade Católica de São Paulo.

Área de Concentração: Linguística Aplicada e Estudos de Linguagem.

Orientador: Prof. Dr. Antonio Paulo Berber Sardinha

**Natalia Luri Arimori Ribeiro**

Portal multimodal/multilíngue para o avanço da Ciência Aberta das Humanidades: Linguagem telecinemática como recurso de ensino de inglês como língua estrangeira através de N-grama de classe semântica.

Aprovado em     /     /     .

Dissertação apresentada à Banca Examinadora da Pontifícia Universidade Católica de São Paulo, como exigência parcial para obtenção do título de Mestre em Linguística Aplicada e Estudos da Linguagem sob orientação do Professor Doutor Antonio Paulo Berber Sardinha

**Banca Examinadora**

---

---

---

Aos meus pais, Lucy Massako Arimori Ribeiro e Pedro Cândido Ribeiro, por sempre me apoiarem e estarem em todos os momentos importantes da minha vida.

À minha batchan, Sueco Arimori, por ter sido a minha maior inspiração como pessoa e o melhor ser humano que já pisou nesta terra.

O presente trabalho foi realizado com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), código de financiamento 130107/2021-2

## AGRADECIMENTOS

Não poderia começar os agradecimentos senão pela minha mãe, Lucy Massako, por sempre ter acreditado em mim e no meu potencial. Sou imensamente grata por ter você como minha mãe e como uma figura inspiradora em minha vida. Você, mãe, junto de meu pai, Pedro Cândido, e do meu irmão, Luís Shizuto, meu nenis, são toda a base de quem eu sou hoje. Seu constante encorajamento, palavras de incentivo e cuidado incansável foram uma fonte constante de força para mim. À minha cunhada, Desiree Lameo, por ser a irmã que eu nunca tive.

Ao meu estimado orientador, Tony Berber Sardinha, gostaria de expressar minha sincera gratidão pelo seu constante apoio, orientação ao longo deste trabalho e confiança nesta “adolescente”, como você mesmo fala. Sua sabedoria, dedicação e paciência foram fundamentais para o meu crescimento acadêmico. Suas ideias inovadoras e perspectivas valiosas abriram caminhos para o meu desenvolvimento como pesquisadora. Sou verdadeiramente grata por ter tido um orientador tão brilhante e humano quanto você.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq, gostaria de expressar minha sincera gratidão por terem financiado minha pesquisa. O apoio financeiro fornecido pelo CNPq foi de extrema importância para o desenvolvimento deste trabalho. Agradeço a confiança e a oportunidade concedidas, contribuindo diretamente para o avanço da ciência em nosso país.

À banca examinadora, Profa. Dra. Sandra Madureira e Profa. Dra. Maria Claudia Delfino, gostaria de expressar minha profunda gratidão por dedicarem seu tempo e expertise na avaliação e discussão do meu trabalho. Seus *insights* valiosos e comentários construtivos foram fundamentais para aprimorar minha pesquisa. Agradeço pela disponibilidade em fazer parte deste processo crucial, contribuindo para a qualidade e relevância do meu trabalho.

À PUC-SP, sou grata por todo o conhecimento adquirido, pelos desafios enfrentados e pelas oportunidades oferecidas para o meu crescimento acadêmico e pessoal. Acredito que essa experiência me preparou de forma abrangente para os desafios futuros e para enfrentar o mundo profissional com confiança.

Marilu, sua atenção, paciência e presteza em ajudar em todas as questões relacionadas ao programa foram inestimáveis. Sempre encontrei em você uma aliada dedicada, pronta para solucionar dúvidas e fornecer informações importantes. Sua presença tornou meu tempo na instituição mais agradável e tranquilo.

Aos meus valorosos colegas do GELC, agradeço por todas contribuições, discussões enriquecedoras e colaboração que foram cruciais para o meu crescimento intelectual e pessoal. Compartilhar ideias e desafios ao longo desta jornada acadêmica com vocês foi uma experiência inestimável. À Me. Aline Zamboni, Profa. Me. Renata Lamberti, Me. Arianne Broginni, mestrandas Quérem Barca e Mirella Whiteman, obrigada por terem me segurado pela mão e dado o apoio, que sem o qual, com certeza não teria chegado onde estou hoje.

Ao Prof. Dr. Leonardo Neves, meu professor da faculdade e amigo pessoal, por ter me apresentado ao LAEL, departamento de Linguística Aplicada e Estudos da Linguagem. Sua indicação e incentivo foram essenciais para que eu pudesse descobrir esse ambiente acadêmico inspirador. Sou grata por sua amizade, pelo excelente professor que foi e por ter me proporcionado a oportunidade de conhecer pessoas incríveis e experiências enriquecedoras.

Gostaria de expressar minha profunda gratidão à Universidade Estadual de Londrina (UEL) por proporcionar uma experiência acadêmica enriquecedora e transformadora. Ao longo do meu tempo na UEL, fui agraciada com professores altamente dedicados e inspiradores, Prof. Dra. Fernanda Brener, Prof. Dra. Adriana Fiori, Prof. Dr. Tacel Leal, Prof. Dr. Elerson Cestaro, Prod. Dr. Pedro Groppo, Prof. Dra. Michelle El Kadri, Prof. Dra. Juliana Tonelli, entre outros, cujo conhecimento e orientação foram essenciais para o meu crescimento intelectual. Sou grata por ter feito parte dessa comunidade acadêmica, que me preparou para os desafios futuros e moldou a pessoa que sou hoje.

Não poderia deixar de agradecer também à minha prima, Daniela Arimori. Ela talvez não saiba o quanto o apoio e acolhimento dela foram cruciais pra eu chegar até aqui. Obrigada por ter me acolhido em sua casa até que eu me estruturasse e conseguisse correr atrás deste sonho.

Aos meus amigos professores que me inspiram na minha profissão, quero expressar minha sincera disposição, pois sempre me inspiraram em minha carreira. Suas paixões pela educação, dedicação aos alunos e comprometimento com a excelência acadêmica são um

verdadeiro exemplo para mim. Agradeço por compartilharem seu conhecimento, experiências e por me encorajarem a alcançar meu potencial como educadora.

Aos meus queridos alunos, expresso profunda gratidão por fazerem parte da minha jornada como educadora. Obrigada por tornarem meu trabalho significativo e por serem uma lembrança constante do impacto positivo que a educação pode ter. Se estou aqui hoje é por vocês e para vocês.

Sei que palavras não são suficientes para expressar completamente minha gratidão, e falar sobre meus sentimentos não é bem a minha qualidade mais a florada, mas espero que saibam que cada um de vocês desempenhou um papel fundamental em minha jornada acadêmica e profissional. Sou imensamente grata por todo o apoio, encorajamento e inspiração que recebi de cada um de vocês. Obrigada do fundo do meu coração.

“A educação é um ato de amor, por isso, um ato de coragem. Não pode temer o debate. A análise da realidade. Não pode fugir à discussão criadora, sob pena de ser uma farsa” (Freire, 2019, p. 127).

## RESUMO

RIBEIRO, N. L. A. **Portal multimodal/multilíngue para o avanço da Ciência Aberta das Humanidades: Linguagem telecinemática como recurso de ensino de inglês como língua estrangeira através de N-grama de classe semântica.** (Dissertação de Mestrado) - LAEL. PUC. São Paulo, 2023

Este estudo focaliza a linguagem telecinemática, uma linguagem verbal utilizada em programas de televisão e filmes de ficção (Bednarek; Veirano Pinto; Werner, 2020). Embora pesquisas anteriores tenham descrito as principais características gramaticais da televisão norte-americana (Berber Sardinha; Veirano Pinto, 2019), britânica (Berber Sardinha; Shimazumi, 2020) e do cinema norte-americano (Veirano Pinto, 2014), nenhuma delas explorou a linguagem telecinemática sob a perspectiva semântica, informando a possibilidade de uso de corpus para o ensino de vocabulário de inglês como língua estrangeira. Esta dissertação preenche essas lacunas ao criar um corpus de linguagem telecinemática e ao informar o ensino de língua estrangeira baseadas em Análise Multidimensional (AMD) e n-grama de classe semântica (NGCS). Adicionalmente, discutimos a implementação de um portal multimodal/multilíngue para ciência aberta nas humanidades, incluindo a curadoria de conteúdos relevantes de dados para Ciências Humanas disponíveis na web e em pesquisas desenvolvidas no LAEL. Este estudo é inovador, pois oferece uma nova abordagem para o ensino de língua estrangeira, utilizando a linguagem telecinemática como ferramenta de ensino, além de propor uma plataforma digital para a ciência aberta nas humanidades.

**Palavras-chave:** Linguagem Telecinemática; N-grama de Classe Semântica; Linguística de Corpus; Análise Multidimensional.

## ABSTRACT

RIBEIRO, N. L. A. **Multimodal/multilingual portal for advancing Open Science in the Humanities: Telecinematic language as a resource for teaching English as a foreign language through Semantic Class N-gram.** (Master's Dissertation) - LAEL. PUC. São Paulo, 2023

This study focuses on telecinematic language, a verbal language used in television programs and fiction films (Bednarek; Veirano Pinto; Werner, 2020). While previous research has described the main grammatical features of American television (Berber Sardinha; Veirano Pinto, 2019), British television (Berber Sardinha; Shimazumi, 2020), and American cinema (Veirano Pinto, 2014), none have explored telecinematic language from a semantic perspective, indicating the potential use of corpora for teaching English vocabulary as a foreign language. This dissertation addresses these gaps by creating a telecinematic language corpus and informing foreign language teaching based on Multidimensional Analysis (MDA) and Semantic Class N-grams (SCNG). Additionally, we discuss the implementation of a multimodal/multilingual portal for open science in the humanities, including curating relevant data content for humanities available on the web and in research conducted at LAEL. This study is innovative as it offers a new approach to foreign language teaching, using telecinematic language as a teaching tool, and proposes a digital platform for open science in the humanities.

**Keywords:** Telecinematic Language; Semantic Class N-gram; Corpus Linguistics; Multidimensional Analysis.

## LISTA DE TABELAS E ILUSTRAÇÕES

FIGURA 1 .....	14
FIGURA 2 .....	16
FIGURA 3 .....	22
FIGURA 4 .....	25
TABELA 1 .....	27
TABELA 2 .....	37
TABELA 3 .....	41

## SUMÁRIO

<b>INTRODUÇÃO</b> .....	1
<b>CAPÍTULO 1 - FUNDAMENTAÇÃO TEÓRICA</b> .....	7
1.1 LINGUÍSTICA DE CORPUS.....	8
1.2 ANÁLISE MULTIDIMENSIONAL (AMD).....	11
1.3 N-GRAMA DE CLASSE SEMÂNTICA (NGCS).....	14
1.4 DISCURSOS TELECINEMÁTICOS COMO FERRAMENTA DE ENSINO.....	17
<b>CAPÍTULO 2 – METODOLOGIA</b> .....	21
2.1 DESIGN, COLETA E LIMPEZA DO CORPUS.....	21
2.2 PROCESSAMENTO DO <i>CORPUS</i> .....	23
<b>CAPÍTULO 3 - ANÁLISE E RESULTADOS</b> .....	26
<b>3.1 ANÁLISE DOS DADOS</b> .....	26
<b>DIMENSÃO 1 - DISCURSO VOLTADO A INSTITUIÇÕES DO GOVERNO VERSUS EXPRESSÃO DE DESEJOS, VONTADES E RELAÇÕES FAMILIARES</b> .....	27
POLO POSITIVO - DISCURSO VOLTADO A INSTITUIÇÕES DO GOVERNO.....	29
POLO NEGATIVO - EXPRESSÃO DE DESEJOS, VONTADES E RELAÇÕES FAMILIARES.....	33
<b>DIMENSÃO 2 – DISCURSO VOLTADO AO TRATAMENTO MÉDICO</b> .....	36
POLO NEGATIVO - DISCURSO VOLTADO AO TRATAMENTO MÉDICO.....	38
<b>DIMENSÃO 3 – DISCURSO DA VIOLÊNCIA VERSUS DISCURSO SENTIMENTAL</b> .....	40
POLO POSITIVO – DISCURSO DA VIOLÊNCIA.....	42
POLO NEGATIVO – DISCURSO SENTIMENTAL.....	44
3.2 DAS DIMENSÕES PARA INFORMAR O ENSINO DE LÍNGUA ESTRANGEIRA.....	46
<b>REFERÊNCIAS</b> .....	53
<b>ANEXOS</b> .....	58

## INTRODUÇÃO

*“I always thought the idea of education was to learn to think for yourself.” (Dead Poets Society)*

A motivação dos alunos nas aulas de língua inglesa em institutos de idiomas e aulas particulares é um desafio crescente, pois muitas vezes as abordagens são banalizadas e focadas excessivamente na tradução e nas estruturas gramaticais de maneira impessoal. Isso leva os alunos a não perceberem a relevância do conteúdo, tornando-se desconectados do que está sendo ensinado. Para contornar esse problema e engajar os alunos em sala de aula, é essencial que eles não apenas acreditem na importância da tarefa, mas também na sua capacidade de realizá-la. Portanto, é importante evitar propor atividades que fujam muito do conhecimento prévio dos alunos, pois isso pode afastá-los ainda mais do aprendizado pretendido.

Uma solução promissora para estimular o engajamento dos alunos é fazer uso da televisão e do cinema, devido ao seu potencial motivador e às oportunidades significativas de exploração didática. A linguagem telecinemática, que compreende a linguagem ficcional presente em filmes, séries e telenovelas, tem sido objeto de estudo sofisticado na literatura acadêmica (Bednarek; Veirano Pinto; Werner, 2020).

Veirano Pinto (2014) oferece uma descrição abrangente da linguagem verbal do cinema, identificando as principais dimensões de variação dos diálogos utilizados na indústria cinematográfica norte-americana desde os anos 1930 até os anos 2000. As dimensões de variação gramaticais identificadas pela autora compreendem os conjuntos de características gramaticais que ocorrem nos textos, ou seja, nos filmes norte-americanos.

No contexto da televisão, há também pesquisas baseadas em corpus sobre a televisão. Berber Sardinha e Veirano Pinto (2019) identificaram as principais dimensões de variação da televisão norte-americana, enquanto Berber Sardinha e Shimazumi (2020) realizaram uma pesquisa semelhante para a televisão britânica. Além disso, Berber Sardinha e Veirano Pinto (2020) apresentaram uma tipologia textual da televisão, fundamentada em corpus, que aponta os principais tipos textuais linguísticos presentes na programação televisiva dos Estados Unidos.

As pesquisas mencionadas são voltadas para identificar a padronização gramatical presentes no cinema e na televisão. No entanto, as pesquisas mencionadas não enfocam a padronização lexical da televisão e do cinema. É essa lacuna que a presente pesquisa visa a preencher.

A presente pesquisa também está direcionada à implementação do 'Portal multimodal/multilíngue para o avanço da ciência aberta nas Humanidades', gestado e desenvolvido no LAEL com o apoio do CNPq. Mais especificamente, o projeto visa a curar conteúdos relevantes de dados para Ciências Humanas disponíveis na web e em pesquisas desenvolvidas no LAEL. Isso inclui a criação e disponibilização de conteúdo inovador no Portal, como um corpus de linguagem telecinemática, e a realização de pesquisas com base em Linguística de Corpus e Análise Multidimensional, voltadas à criação de atividades de ensino de inglês com foco comunicativo, que também serão disponibilizadas no Portal, levando em conta as dimensões de variação da linguagem telecinemática.

O Portal proposto pelo LAEL é sustentado por quatro grandes pilares, a saber: 'acesso aberto (*open access*)', 'dados abertos (*open data*)', 'fonte aberta (*open source*)', e 'ciência cidadã (*citizen science*)' (Berber Sardinha et al, 2021, p.4). Esses fundamentos são essenciais para promover a disseminação do conhecimento e a participação ativa da comunidade científica e interessados na área de Ciências Humanas. Segundo os autores:

O acesso aberto prevê o acesso imediato e gratuito ao usuário final à comunicação científica. A política de dados abertos preconiza o acesso gratuito e descomplicado a dados de pesquisa a todos, sem restrição de direitos autorais ou outros mecanismos restritivos. A fonte aberta permite que o código de programas de computador, scripts e aplicativos seja compartilhado por meio de licenças que prevejam a possibilidade de o código ser alterado por outros usuários. E por fim, a ciência cidadã estimula o engajamento do público na pesquisa científica, não somente os especialistas. (Berber Sardinha et al, 2021, p.4),

Uma das motivações fundamentais para a criação do Portal é a necessidade urgente de promover uma mudança de paradigma nas Ciências Humanas, que ainda são caracterizadas por práticas fechadas e herméticas. Embora a Ciência Aberta já tenha ganhado ampla aceitação em outras áreas do conhecimento, nas Ciências Humanas, essa abordagem ainda é uma exceção (Berber Sardinha et al, 2021). Especificamente na Linguística e Linguística Aplicada, é raro

que os pesquisadores compartilhem seus dados, protocolos e software com a comunidade acadêmica.

Outra particularidade é a tendência dos pesquisadores em enxergar-se como autossuficientes, assumindo a responsabilidade por todas as etapas da pesquisa, desde o seu desenvolvimento até a implementação e divulgação. Em contraste, nas áreas das Ciências da Saúde, Engenharias e Computação, por exemplo, observa-se uma maior aderência às práticas da ciência aberta (Berber Sardinha et al, 2021), onde os pesquisadores trabalham frequentemente em equipes amplas, distribuindo tarefas entre grupos especializados e priorizando a divulgação ágil dos resultados.

Essa diferença de abordagem tem sérias consequências, incluindo a demora na disseminação do conhecimento científico, uma menor produtividade dos pesquisadores e uma participação limitada das Ciências Humanas no debate público sobre ciência. Essa falta de compartilhamento e colaboração pode impedir o avanço do conhecimento, além de restringir o impacto das pesquisas desenvolvidas nessas áreas.

Portanto, o Portal Multimodal/Multilíngue para o Avanço da Ciência Aberta nas Humanidades tem como objetivo enfrentar esses desafios e promover uma mudança significativa no cenário das Ciências Humanas. Ao incentivar a prática da ciência aberta e disponibilizar conteúdos relevantes de dados, protocolos e software, o Portal busca promover a colaboração entre pesquisadores, estimular o debate e ampliar a visibilidade das pesquisas desenvolvidas nas áreas de Linguística e Linguística Aplicada.

Através desse novo paradigma, espera-se alcançar resultados significativos, acelerando a comunicação da ciência, impulsionando a produtividade dos pesquisadores e fortalecendo a participação das Ciências Humanas no contexto científico e na sociedade como um todo. A promoção de uma cultura de abertura e colaboração, tão essencial no Portal, contribuirá para impulsionar o progresso e a relevância dessas áreas no cenário da ciência aberta.

Vale destacar ainda que o Portal, ao disponibilizar essas informações, se tornará o primeiro meio de divulgação de pesquisa de variação lexical da televisão e do cinema, o que amplificará sua relevância e influência. A relevância e o pioneirismo também se evidenciam ao considerar que esta é a primeira dissertação de mestrado a utilizar essa metodologia. Isso

demonstra o quão inovador e valioso é o trabalho realizado, abrindo novos caminhos para o avanço das Ciências Humanas e sua integração no movimento da Ciência Aberta.

Com a implementação bem-sucedida dessa metodologia e sua divulgação por meio do Portal, espera-se que mais pesquisadores se engajem nessa abordagem colaborativa, contribuindo para uma Ciência Aberta nas Ciências Humanas, mais dinâmica, participativa e alinhada com as práticas adotadas em outras áreas do conhecimento. O impacto dessa mudança de paradigma pode ser significativo, ampliando o alcance das pesquisas, a troca de conhecimentos e a importância das Ciências Humanas no cenário científico e na sociedade em geral.

O presente projeto assume uma importância significativa ao visar a preencher três lacunas na literatura. A primeira delas refere-se à falta de uma descrição discursiva baseada em corpus da linguagem telecinemática, que abrange tanto a televisão quanto o cinema. Embora já existam descrições discursivas baseadas em corpus para outros domínios, como a escrita acadêmica e a literatura de ficção, ainda não há uma descrição detalhada e sofisticada dessa forma de linguagem utilizada em produções televisivas e cinematográficas.

A segunda lacuna que este estudo busca preencher está relacionada à ausência de propostas de utilização das dimensões de variação lexical no ensino de línguas, em especial no contexto do ensino de inglês como língua estrangeira. Embora as dimensões de variação tenham sido identificadas em pesquisas anteriores, abrangendo o cinema e a televisão norte-americana e britânica, ainda não foram devidamente aplicadas em sala de aula com alunos de inglês como língua estrangeira.

Por fim, a terceira lacuna a ser preenchida é a disponibilização de um corpus de linguagem telecinemática por meio do Portal. Esse corpus será uma contribuição inédita para a comunidade acadêmica, permitindo uma descrição abrangente da linguagem telecinemática em inglês com enfoque lexical-discursivo. Com base nessa descrição, o projeto irá propor atividades de ensino de inglês para institutos de idiomas e aulas particulares, explorando as dimensões de variação junto aos alunos. Essas atividades serão disponibilizadas no Portal, garantindo acesso aberto a todos interessados.

Portanto, a iniciativa visa a preencher as lacunas existentes na literatura ao oferecer uma descrição lexical detalhada da linguagem telecinemática, informar as dimensões de variação no

ensino de inglês e ao disponibilizar um corpus inédito para a comunidade acadêmica. Além disso, o projeto pretende enriquecer o ensino de inglês como língua estrangeira, permitindo que os alunos explorem plenamente o potencial dessa forma de linguagem, promovendo assim uma maior eficácia e engajamento nas aulas. Ao divulgar e compartilhar esses recursos por meio do Portal, espera-se contribuir significativamente para a evolução das Ciências Humanas e para a disseminação da ciência aberta nesse campo específico.

A partir dessas lacunas, identificamos três objetivos específicos para a presente pesquisa:

1. Descrever a linguagem telecinemática de maneira discursiva, através das dimensões de N-Gramas de Classes Semânticas.
2. Informar o ensino de inglês como língua estrangeira baseadas nas dimensões de variação.
3. Alimentar o Portal das Humanidades com um corpus de linguagem telecinemática.

Com esses três objetivos em mente, esta pesquisa visa a responder às seguintes perguntas:

1. Quais as dimensões lexicais de N-Gramas de Classes Semânticas presentes no discurso telecinemático?
2. Como as dimensões identificadas podem informar o ensino de vocabulário de inglês como língua estrangeira?

O trabalho será estruturado em quatro partes principais, cada uma contribuindo para alcançar o propósito geral da pesquisa:

1. **Fundamentação Teórica:** No primeiro capítulo, serão abordadas as bases teóricas essenciais, incluindo as abordagens metodológicas da Linguística de Corpus e da Análise Multidimensional, juntamente com o uso das N-Gramas de Classe Semânticas. Além disso, será apresentada uma explanação sobre a importância dos discursos telecinemáticos no contexto do ensino da língua inglesa. Essa seção fornecerá uma base sólida para o desenvolvimento da pesquisa.
2. **Metodologia:** O segundo capítulo detalhará a metodologia adotada para a pesquisa, incluindo o processo de coleta e limpeza de texto, bem como os procedimentos automáticos de processamento e etiquetagem dos dados. Será descrito o passo a passo para obter os

dados necessários, bem como as técnicas empregadas para facilitar a análise e identificação de padrões discursivos relevantes.

3. Análise das Dimensões e Proposta para Informar o Ensino de Língua: O terceiro capítulo, será dividido em duas seções. A primeira realizará a análise das dimensões identificadas na fase anterior, levando em consideração os resultados obtidos. Essa análise contribuirá para uma compreensão aprofundada das características da linguagem telecinemática em inglês, permitindo a identificação de padrões e tendências significativas. A segunda, consistirá na apresentação de uma proposta para informar o ensino de língua inglesa através das dimensões encontradas, visando proporcionar atividades práticas para os estudantes aproveitarem as características e aspectos relevantes dos discursos telecinemáticos.

Na última parte deste trabalho, as considerações finais serão apresentadas. Haverá uma síntese dos principais pontos discutidos ao longo da pesquisa, enfatizando as descobertas significativas resultantes das análises de dados. Serão feitas reflexões sobre a efetividade das propostas didáticas e suas implicações para o ensino da língua inglesa. Além disso, serão apontadas sugestões para futuras pesquisas e aprimoramentos na metodologia utilizada, destacando a relevância do estudo e suas possíveis contribuições para o avanço do campo acadêmico.

## CAPÍTULO 1 - FUNDAMENTAÇÃO TEÓRICA

*“Do, or do not. There is no try.” (Star Wars)*

A fundamentação teórica do presente projeto é composta por quatro etapas essenciais, cada uma delas fornecendo a base conceitual e metodológica para a pesquisa:

Na primeira etapa, será realizada uma exposição acerca da Linguística de Corpus, abordando seus princípios e conceitos fundamentais. Serão explorados aspectos relacionados à coleta, organização e análise de grandes volumes de dados textuais. A LC é uma abordagem valiosa para investigar padrões linguísticos presentes em diferentes contextos e gêneros textuais, fornecendo insights importantes para a compreensão da linguagem telecinemática.

A segunda etapa abordará a Análise Multidimensional, a vertente da LC empregada nesta pesquisa. Serão apresentados os fundamentos teóricos dessa abordagem, juntamente com as técnicas empregadas para analisar múltiplas dimensões dos dados coletados. A AMD proporciona uma visão abrangente das características linguísticas presentes nos textos analisados, permitindo uma investigação mais profunda e detalhada dos discursos telecinemáticos. Também serão apresentados outros estudos científicos que utilizaram com sucesso a AMD.

A terceira etapa é dedicada ao conceito de N-Grama de Classe Semântica, uma abordagem inovadora no âmbito da AMD que descreve unidades linguísticas formadas por sequências de palavras e sua relação com classes semânticas específicas. Será explorada a utilização dessa abordagem para identificar padrões de linguagem dentro dos discursos telecinemáticos. Essa análise detalhada contribuirá para a compreensão mais aprofundada da linguagem empregada em produções televisivas e cinematográficas.

A quarta e última etapa destaca a linguagem telecinemática como uma ferramenta eficaz no ensino de inglês como língua estrangeira. Serão discutidos os benefícios e as possibilidades de utilizar discursos telecinemáticos no contexto educacional, enfatizando sua relevância para o desenvolvimento das habilidades linguísticas dos estudantes. Essa etapa demonstrará como a análise e o entendimento da linguagem telecinemática podem ser aplicados para criar atividades e recursos que engajem os alunos e tornem o processo de aprendizado mais envolvente e efetivo.

Através dessas quatro etapas, o projeto se apoia em sólidas bases teóricas e metodológicas, proporcionando uma abordagem abrangente e inovadora para explorar a linguagem telecinemática e suas aplicações no ensino de inglês como língua estrangeira. Ao preencher essas lacunas na literatura e oferecer novos insights, espera-se contribuir significativamente para o avanço das Ciências Humanas e para o aprimoramento das práticas educacionais.

## 1.1 LINGUÍSTICA DE CORPUS

A Linguística de *Corpus* (LC) é a base teórica fundamental para esta pesquisa, sendo uma abordagem que se concentra na coleta e exploração de corpora, ou seja, conjuntos de dados linguísticos textuais coletados de forma criteriosa para fins de pesquisa sobre uma língua ou variedade linguística específica (Berber Sardinha, 2000, p.3). O objetivo principal da LC é investigar a língua em seu uso real, refletindo a forma como as pessoas realmente falam e utilizam a linguagem no cotidiano (Biber; Conrad; Reppen, 1998).

Conforme destacado por Delfino (2016), a LC surgiu da necessidade percebida pelos linguistas de se basearem na linguagem utilizada no dia a dia das pessoas para realizar generalizações e desenvolver teorias sobre o funcionamento da língua. Isso significa que a análise é realizada a partir da observação da língua em contexto real, em situações autênticas de uso.

Sinclair (2004, p. 52) descreve a língua em uso como "uma negociação contínua entre os participantes e, ao mesmo tempo, um registro em desenvolvimento de experiências."<sup>1</sup> Essa abordagem reforça a visão da LC como uma forma de compreender a língua em sua dinamicidade e interação com os falantes, em contraste com uma análise estática e descontextualizada.

---

<sup>1</sup> Original: "both a continuous negotiation between participants, and a developing record of experience." (Sinclair, 2004, p.52)

É importante ressaltar que a LC não é considerada uma metodologia específica, mas sim uma abordagem baseada em *corpus* (Gil, 2017, apud Biber; Conrad; Reppen, 1998). Em vez de ser uma metodologia rígida e prescritiva, a LC é flexível e se adapta a diferentes contextos e objetivos de pesquisa. Primeiramente, é necessário definir o que é *corpus*. Berber Sardinha (2000) discorre que a definição de um *corpus* deve levar em consideração:

- (a) A origem: Os dados devem ser autênticos;
- (b) O propósito: O corpus deve ter a finalidade de ser um objeto de estudo linguístico;
- (c) A composição: O conteúdo do corpus deve ser criteriosamente escolhido;
- (d) A formatação: Os dados do corpus devem ser legíveis por computador;
- (e) A representatividade: O corpus deve ser representativo de uma língua ou variedade;
- (f) A extensão: O corpus deve ser vasto para ser representativo (2000, p. 336)

Nesta pesquisa, para estabelecer os padrões da linguagem cinematográfica, serão utilizados textos retirados das próprias obras cinematográficas, ou seja, materiais autênticos provenientes dos filmes. A escolha por materiais autênticos é fundamental, pois eles refletem a linguagem real utilizada no contexto cinematográfico, com suas características específicas e autênticas. Dessa forma, os discursos telecinemáticos serão analisados em seu contexto original, o que proporcionará uma compreensão mais precisa e fiel da linguagem presente nessas produções.

Para a construção do corpus, é essencial que sejam definidos requisitos específicos para a seleção dos filmes e textos que o comporão. O número de filmes escolhidos deve ser suficientemente amplo para garantir uma análise abrangente, levando em conta a diversidade de produções cinematográficas e suas características linguísticas distintas.

No entanto, o conceito de "representatividade" é considerado controverso, conforme pontuado por Berber Sardinha (2000). Ao indagarmos sobre o que é representativo e para quem, percebemos que a representatividade absoluta é impraticável, especialmente quando se trata de descrever padrões de linguagem da língua inglesa, pois não é possível obter uma amostragem de toda a população falante desse idioma.

Em vez disso, é mais adequado considerar a representatividade de maneira relativa. A linguagem é caracterizada por sua natureza probabilística, ou seja, existem traços linguísticos

mais comuns e menos comuns em determinados contextos. Portanto, a análise do corpus permitirá estabelecer relações entre esses traços, identificando padrões que são mais frequentes em produções cinematográficas em inglês.

Quanto à segunda pergunta, os criadores do corpus devem ter cautela ao fazer generalizações dos resultados obtidos e atribuí-los como representativos de um segmento específico. É importante reconhecer que o corpus selecionado pode representar um conjunto específico de filmes e gêneros cinematográficos, mas não pode ser considerado representativo de todas as produções em inglês. Portanto, a interpretação dos resultados deve ser feita com a devida consideração do escopo do corpus e do contexto em que foi coletado.

Para que a análise baseada em corpus seja possível, Biber (1998, p. 4) estabelece quatro características:

1. Ser empírica, devendo analisar os padrões da língua em seu uso natural;
2. Deve utilizar uma grande quantidade de textos, chamados de “corpus”;
3. Haver computadores para que a análise seja feita;
4. Utilizar-se técnicas quantitativas e qualitativas.

O uso de computadores é fundamental na análise da Linguística de Corpus (LC), uma vez que o número extenso de textos linguísticos coletados para investigar os padrões de linguagem torna inviável a análise manual a olho nu. Os computadores têm a vantagem de não se cansarem como os humanos, possibilitando uma análise consistente e detalhada dos dados. Essa abordagem permite um mapeamento mais preciso e abrangente da linguagem em uso na sociedade, fornecendo insights sem precedentes sobre as características linguísticas e seu uso e manipulação na comunicação cotidiana (Araújo, 2017, apud McEnery, 2013).

A LC trata a língua do ponto de vista do uso, ou seja, das escolhas linguísticas não aleatórias, feitas por falantes em situações reais do cotidiano, que seguem padrões específicos. Com a LC, é possível analisar e quantificar os padrões e variações linguísticas, estabelecendo correlações entre os traços linguísticos e os contextos situacionais de uso da linguagem (Delfino, 2022, p. 22). Por meio da coleta criteriosa e da análise computacional dos dados, é possível desenvolver pesquisas sobre uma língua ou variedade linguística específica,

possibilitando uma compreensão mais profunda e abrangente da linguagem em uso na sociedade (Berber Sardinha, 2000, p. 325).

## 1.2 ANÁLISE MULTIDIMENSIONAL (AMD)

A Análise Multidimensional (AMD), criada por Douglas Biber, será a base metodológica desta pesquisa. Essa abordagem oferece uma descrição abrangente e complexa de corpora inteiros de textos, utilizando métodos estatísticos para identificar características textuais comuns entre eles (Berber Sardinha, 2000, p. 101). O objetivo principal da AMD é identificar coocorrências linguísticas tanto quantitativa quanto qualitativamente, comparando registros de linguagem falada e/ou escrita (Kauffmann, 2020, p. 33).

O nome "Análise Multidimensional" reflete a capacidade dessa abordagem de reconhecer e comparar diferentes dimensões de variação dentro de um espaço linguístico (Romeiro, 2020, p. 24). Biber (1988, p. 55) postula que essas coocorrências não são aleatórias, mas sim, possuem uma função em comum, ou seja, um padrão. Esses padrões podem se referir a repetições observáveis de características linguísticas, tanto em termos de estruturas gramaticais quanto lexicais (Trigo, 2020, p. 91, apud Hunston, 2010).

A AMD compreende duas vertentes principais: a Funcional e a Lexical. A AMD Funcional, proposta por Biber (1988), tem como objetivo identificar padrões gramaticais da língua. Em seu estudo seminal, Biber analisou 481 textos de 23 gêneros textuais diferentes, totalizando aproximadamente 960 mil palavras. Cada texto componente é considerado individualmente, e todas as variáveis são analisadas simultaneamente (Brogini, 2023, p. 19). Inicialmente, a análise multifatorial e multidimensional identificou sete dimensões, mas a sétima não possuía variáveis suficientes para ser considerada neste primeiro estudo, requerendo investigações futuras mais aprofundadas. Dito isso, as seis dimensões analisadas foram (Biber, 1988):

1. Produção interacional x produção informacional;
2. Propósitos narrativos x não narrativos;

3. Referência explícita x dependente da situação;
4. Persuasão explícita;
5. Informação abstrata x não abstrata;
6. Elaboração em tempo real.

A Análise Multidimensional Lexical, que também é uma das bases para este estudo, difere da AMD Funcional em seu objetivo e abordagem. Enquanto a AMD Funcional busca reconhecer coocorrências de estruturas gramaticais em um número específico de categorias gramaticais, a AMD Lexical tem como propósito revelar discursos subjacentes à superfície linguística de muitos textos, sob uma análise multivariada (Brogini, 2023, p. 19).

A AMD Lexical visa identificar a variação do discurso de acordo com o contexto, sem restrições a um número limitado de variáveis. Enquanto a AMD Funcional foca nas estruturas gramaticais, a AMD Lexical se concentra em extrair e analisar fatores relacionados ao uso de palavras e expressões específicas que são significativas para a compreensão dos discursos subjacentes presentes nos textos (Kauffmann, 2020, p. 44).

A realização da análise com base na AMD Lexical envolve os seguintes passos, conforme descritos por Brogini (2023, p. 21, apud Berber Sardinha, 2014, 2017):

1. Identificação e contagem das palavras;
2. Normalizar das frequências das variáveis lexicais;
3. Extração fatorial inicial não rotacionada baseada nas frequências normalizadas para identificar os fatores a serem utilizados;
4. Scree plot: Definição do número de fatores para análise por meio de um gráfico de análise de sedimentação;
5. Eliminação das variáveis lexicais com comunalidades menores que 0,15 (cf. Biber, 2006, p. 183);
6. Extração fatorial final rotacionada contendo o número de fatores estabelecidos para a análise;
7. Cálculo da quantidade de variação compartilhada pelos fatores extraídos;
8. Checagem da variância dos fatores;
9. Cálculo dos escores de fator de cada texto;
10. Interpretação dos fatores em termos de seus discursos subjacentes por meio da observação dos textos, registros e variáveis.

Esses passos permitem analisar os discursos subjacentes presentes nos textos, identificando padrões e variações lexicais que refletem os contextos específicos em que a

linguagem é utilizada nos discursos telecinemáticos. A AMD Lexical proporciona uma abordagem abrangente e detalhada para compreender as características linguísticas dessas produções cinematográficas em inglês e como elas contribuem para a construção do sentido nos filmes.

De fato, a Análise Multidimensional (AMD), tanto na sua vertente Funcional quanto na Lexical, tem sido um arcabouço teórico valioso para diversas pesquisas em diferentes contextos linguísticos e discursivos. Ela tem sido aplicada em uma ampla gama de estudos, abrangendo desde a análise da linguagem telecinemática, como as pesquisas de Berber Sardinha e Veirano Pinto (2019) sobre as variações na linguagem da TV estadunidense e de Berber Sardinha e Shimazumi (2020) sobre a linguagem da TV britânica, até outras investigações em áreas diversas.

Por exemplo, Brogini (2023) realizou uma pesquisa sobre as representações da palavra "sustentabilidade" nas redes sociais, aplicando a AMD Lexical para revelar os discursos subjacentes associados ao termo nesse contexto específico. Delfino (2022) empreendeu uma análise multidimensional e multimodal das músicas de língua inglesa, utilizando a AMD para compreender a variação linguística presente nessas produções musicais. Romeiro (2020) explorou o discurso verbal nas artes, especificamente nas fotografias de Sally Mann, por meio da aplicação da AMD, que permitiu identificar padrões e variações na linguagem presentes nessas obras de arte. Kauffmann (2020), por sua vez, utilizou toda a obra de Machado de Assis como *corpus* para analisar a variação linguística do autor e compreender suas características estilísticas.

Em resumo, a AMD é uma ferramenta analítica poderosa que pode ser aplicada a diferentes domínios da língua em uso. Por ser multidimensional, possibilita uma compreensão mais profunda das características linguísticas e discursivas em uma variedade de contextos, permitindo que os pesquisadores desvendem padrões, tendências e variações significativas na linguagem utilizada em diferentes formas de expressão, como filmes, redes sociais, músicas, artes visuais e literatura. Nesta pesquisa, continuamos a expandir os limites da AMD, aplicando-a à análise de n-gramas de classe semântica, conforme descrito a seguir.

### 1.3 N-GRAMA DE CLASSE SEMÂNTICA (NGCS)

A aplicação da metodologia de N-grama de Classe Semântica (NGCS; Berber Sardinha, 2023) nesta pesquisa representa um avanço significativo na análise dos discursos telecinemáticos em inglês. Os NGCS são sequências de três classes semânticas, que ocorrem nos textos. Essas classes semânticas são definidas pela tipologia semântica do USAS, UCREL Semantic Analysis System (Rayson, 2007). Por exemplo, o NGCS “Z5\_Z5\_B3” composto por dois *Grammatical bins* (Z5), que representam preposições, advérbios ou conjunções, e um *Medical* (B3), que engloba termos relacionados com medicação ou tratamentos médicos.

A pesquisa tem como propósito identificar os NGCSs mais frequentes no corpus telecinemático em inglês, bem como reconhecer os discursos subjacentes ao uso dessas combinações de palavras. Além disso, será investigada a possibilidade de utilizar os NGCS como uma ferramenta de ensino da língua inglesa, visando aprimorar o processo de aprendizado e compreensão da língua em um contexto comunicativo autêntico e relevante.

Conforme colocado, para realizar a identificação dos agrupamentos semânticos, a pesquisa se apoia no Sistema de Análise Semântica UCREL (UCREL Semantic Analysis System), ou USAS, desenvolvido pela Universidade de Lancaster. O USAS classifica o léxico em 21 categorias maiores, denominadas campos semânticos, representadas por letras do alfabeto. Essa abordagem permite explorar os padrões de uso da linguagem e compreender as nuances semânticas presentes nos discursos telecinemáticos analisados.

Figura 1

Código	Categoria Semântica (Português e Inglês)	
A	Termos genéricos e abstratos	General and abstract terms
B	O corpo e o indivíduo	The body and the individual
C	Artes e artesanatos	Arts and crafts
E	Emoções	Emotions
F	Comida e agricultura	Food and farming
G	Governo e domínio público	Government and public

H	Arquitetura, prédios, casas e lar	Architecture, housing and the home
I	Dinheiro e comércio na indústria	Money and commerce in industry
K	Entretenimento, esportes e jogos	Entertainment, sports and games
L	Vida e seres vivos	Life and living things
M	Movimento, localização, viagem e transporte	Movimento, location, travel and transport
N	Números e medidas	Numbers and measurement
O	Substâncias, materiais, objetos e equipamentos	Substances, materials, objects and equipment
P	Educação	Education
Q	Linguagem e comunicação	Language and communication
S	Ações, estados e processos sociais	Social actions, states and processes
T	Tempo	Time
W	Mundo e meio ambiente	World and environment
X	Ações, estados e processos psicológicos	Psychological actions, states and processes
Y	Ciência e tecnologia	Science and technology
Z	Nomes próprios e gramática	Names and grammar

Fonte: Ribeiro (2023) para esta pesquisa, adaptado de Archer et. al., p. 2, 2004.

O USAS é um sistema de classificação lexical que categoriza palavras em campos semânticos específicos. Cada campo geral de discurso é representado por uma letra do alfabeto (por exemplo, A para termos gerais e abstratos). Essas letras são acompanhadas de subcategorias representadas por dígitos. Caso haja mais de uma subcategoria, são adicionados dígitos decimais, como podemos ver na figura abaixo:

Figura 2

A1	General and abstract terms
A1.1.1	General actions, making etc.
A1.1.2	Damaging and destroying
A1.2	Suitability
A1.3	Caution
A1.4	Chance, luck
A1.5	Use
A1.5.1	Using
A1.5.2	Usefulness
A1.6	Physical/mental
A1.7	Constraint
A1.8	Inclusion/Exclusion
A1.9	Avoiding
A2	Affect
A2.1	Affect:- Modify, change
A2.2	Affect:- Cause/Connected
A3	Being

Fonte: Ribeiro (2023) para esta pesquisa.<sup>2</sup>

Os NGCSs assemelham-se, na superfície, aos *bundles* lexicais propostos por Viviana Cortes (2004). Os *bundles* lexicais são conjuntos de três ou mais palavras que frequentemente ultrapassam as fronteiras gramaticais e têm natureza altamente formulaica. Essas combinações de palavras coocorrem com frequência dentro de um corpus específico. Exemplos de *bundles* lexicais incluem expressões como "*at the end of the day*", "*as a matter of fact*", "*in other words*" e "*on the other hand*".

No entanto, a diferença central entre *bundles* lexicais e NGCS reside no objeto de estudo. Enquanto os *bundles* lexicais se concentram em identificar padrões de combinações de palavras para desempenhar funções gramaticais específicas, os NGCS buscam identificar combinações recorrentes lexicais dentro de discursos particulares.

---

<sup>2</sup> Tabela completa no apêndice.

## 1.4 DISCURSOS TELECINEMÁTICOS COMO FERRAMENTA DE ENSINO

O estudo da linguística telecinemática é uma área de pesquisa recente, que pertence à cultura pop (Bednarek, 2018). Por muito tempo, essa linguagem não foi reconhecida como um objeto relevante de estudo, sendo parte desse preconceito causado pelo fato de que as falas dos personagens nos filmes eram criadas, muitas vezes, por motivos artísticos e não refletiam necessariamente o contexto social (Veirano Pinto, 2014, p. 110 apud Alvarez-Pereyre, 2011, p. 57).

O cinema é muito mais do que uma simples fonte de entretenimento. De acordo com Araújo e Voss (2009), ele sempre teve a capacidade persuasiva, uma vez que representa a realidade. Na pesquisa de Quinta Junior (2019) sobre o trabalho de Leni Riefenstahl para o governo nazista de Hitler, é discutido como a escolha estética da diretora na obra "O Triunfo da Vontade" (1935) serviu como uma forma de propaganda indireta do *Terceiro Reich*, sendo elaborada de maneira que a audiência não percebesse a intencionalidade.

Além de representar o mundo, o cinema pode ter um impacto real na realidade. Em 2013, após o lançamento do filme "Lincoln", de Steven Spielberg que mostrava a luta do então presidente para conseguir os votos necessários para a aprovação da 13ª Emenda que abolia a escravatura nos Estados Unidos, Ranjan Batra e Ken Sullivan ficaram intrigados sobre o processo de ratificação da emenda no estado de Mississippi. Em 1995, começava um processo de ratificação da emenda que aboliu a escravidão em todo o território nacional, mas nenhum dos dois lembraram do resultado e, após pesquisas, descobriram que ela nunca havia sido totalmente ratificada pois, após aprovação no Congresso estadual, deveria ter sido enviada ao Secretário de Estado para assinatura, o que nunca aconteceu. Com isso, após Sullivan contatar o Secretário de Estado de Mississippi, Delbert Hosemann, a emenda foi finalmente assinada, 147 anos após ela ter sido oficialmente anexada à Constituição dos Estados Unidos da América (Blum, 2013).

O cinema pode ainda ser considerado uma reconstrução da sociedade e uma prática social que desperta o fascínio dos espectadores (Holleben, 2007). Como ferramenta educativa, o cinema oferece uma oportunidade única de ampliar o ensino para fora da sala de aula,

permitindo que os alunos interpretem as obras de acordo com suas perspectivas individuais e conhecimento de mundo (Fabris, 2008). Segundo a autora,

Analisar uma produção como o cinema, que rompe com as formas mais comuns de representação, em que a materialidade é a imagem em movimento, é ingressar em uma outra dimensão do conhecimento. A imagem em movimento não apenas tenta reproduzir o “real”, como também nos faz entrar em uma dimensão espaço-temporal singular, criando um jeito novo de conhecer através do olho da câmera; nesse processo, pensamos e conhecemos por imagens (2008, p 125)

Através desse processo, os alunos são capazes de sentir que fazem parte da história e, conseqüentemente, do próprio processo de conhecimento. Além disso, o desenvolvimento do cinema e do audiovisual trouxe consigo a popularização das séries de TV, que se tornaram amplamente acessíveis por meio de serviços de streaming. Essas séries oferecem uma alternativa mais concisa e acessível em comparação aos filmes, permitindo que os alunos assistam várias vezes e aprofundem sua compreensão (Franco, 2018). Essa facilidade de acesso ao conteúdo proporciona aos estudantes a oportunidade de se envolverem mais profundamente com as narrativas, enredos e personagens, enriquecendo sua experiência de aprendizado através do audiovisual.

No entanto, para que o uso dessas mídias seja efetivo em sala de aula, é essencial que os professores estejam preparados para a intervenção. Ferreira (2009) enfatiza a importância de os professores estarem conscientes do uso desse recurso como objeto de investigação, suporte de sensibilização, base para questionamento ou recurso pedagógico. Ao escolherem utilizar discursos telecinemáticos na educação, os educadores reconhecem a relevância do conhecimento de mundo dos alunos. No entanto, é fundamental que toda atividade baseada nessas mídias visuais tenha um objetivo pré-estabelecido, para que vá além do mero entretenimento. Nesse contexto, os professores desempenham um papel importante, não apenas na condução da intervenção, mas também na compreensão e identificação de seus alunos.

Outro aspecto relevante a ser considerado é que as escolas precisam preparar os alunos para um mundo em constante mudança, deixando de lado recursos "arcaicos" (Ferreira, 2009). A incorporação de novas mídias reconhece outras fontes de conhecimento, ampliando o papel da língua como uma forma de acessar novos aprendizados além do ensino da língua em si. Ela passa a ser utilizada como uma fonte constante de representação para alcançar propósitos

culturais diversos (The New London Group, 1996). Acredita-se que o uso do cinema no processo de aprendizagem pode facilitar a aproximação e identificação dos alunos com a realidade, despertando maior interesse pelo objeto de estudo.

A abordagem comunicativa é uma das bases para a intervenção proposta, buscando a integração equilibrada das quatro habilidades linguísticas: fala, escrita, audição e leitura (Leffa, 1988). Brown (2007) destaca a importância de conectar os aspectos formais da linguagem, como gramática e discurso, aos aspectos pragmáticos, como funcionalidade, sociolinguística e estratégias comunicativas. O foco é engajar os alunos no uso autêntico e pragmático da linguagem com propósitos significativos, em que as formas organizacionais da linguagem são consideradas como componentes importantes para alcançar esses propósitos.

Nesse contexto, o papel do aluno é de agente ativo em seu próprio processo de aprendizado. O professor assume o papel de orientador e facilitador, em vez de ser a única autoridade detentora do conhecimento em sala de aula (Leffa, 1988). A ênfase está em encorajar os alunos a criarem seus próprios significados e envolvê-los em interações reais. Para isso, é fundamental utilizar tecnologias, como vídeos, para permitir que o aprendizado aconteça de forma mais espontânea e autêntica (Brown, 2007). A utilização de materiais autênticos é essencial para garantir a realidade das situações e evitar simplificações que poderiam interferir na autenticidade do aprendizado (Leffa, 1998).

O uso de discursos telecinemáticos pode acelerar o processo de aprendizagem dos alunos, considerando os pilares da aprendizagem propostos por Dehaene (2002): atenção, envolvimento ativo, feedback e consolidação. Capturar a atenção dos alunos é crucial para a aprendizagem, uma vez que a atenção consciente permite que as informações sejam assimiladas e processadas de forma mais profunda e eficaz (Dehaene, 2002). Nesse sentido, filmes e séries podem ser ferramentas poderosas para direcionar e manter a atenção dos alunos, potencializando sua aprendizagem.

As atividades propostas devem evitar que os alunos se tornem meros espectadores passivos no processo de aprendizado. Segundo Dehaene (2022), o envolvimento ativo dos alunos não se limita a movimentos físicos, mas sim a participação ativa e engajada em compreender e interagir com a língua, buscando entender sentenças em língua estrangeira sem depender de terceiros (p. 246). No entanto, é fundamental que os alunos não sejam deixados "à

própria sorte" (p. 252). Cabe ao professor estar preparado para orientar, ter uma pedagogia clara e objetivos definidos para engajar a curiosidade do aluno. A curiosidade é um fator importante para o aprendizado, pois impulsiona a busca por conhecimento (Dehaene, 2022, p. 263).

As propostas desta pesquisa serão fundamentadas na abordagem proposta por Brown (2007). Segundo o autor, as técnicas de linguagem devem ser elaboradas com o objetivo de envolver os alunos no uso pragmático, autêntico e funcional da língua para propósitos significativos. Nessa perspectiva, o foco principal não está nas estruturas gramaticais isoladas, mas sim na aplicação real e contextualizada da língua, permitindo que o aprendiz alcance seus objetivos comunicativos (Brown, 2007, p. 46).

Essa abordagem valoriza o papel central do aluno na aprendizagem, considerando-os participantes ativos em seu próprio processo de conhecimento. Os alunos são encorajados a construir significados através de interações reais e práticas com a língua, o que proporciona uma aprendizagem mais significativa e duradoura. Nesse contexto, o professor atua como facilitador e guia, dando espaço para que os alunos se expressem e interajam com a língua em situações autênticas. O objetivo é prepará-los para utilizar a língua de forma produtiva e receptiva em contextos reais fora da sala de aula (Brown, 2007, p. 47). A utilização de tecnologia, como filmes e programas de televisão, é considerada uma ferramenta valiosa para a abordagem centrada no aluno. Esses recursos proporcionam aos estudantes oportunidades de exposição à língua em contextos autênticos e variados, o que contribui para o desenvolvimento de habilidades comunicativas e para a compreensão de nuances linguísticas presentes em situações reais.

## CAPÍTULO 2 – METODOLOGIA

*“Life only really has one beginning and one end, and the rest is just a whole lot of middle.” (Glee)*

Conforme colocado, a metodologia desta pesquisa baseia-se na análise de discursos telecinemáticos, utilizando uma combinação da abordagem da Análise Multidimensional Lexical (AMD) com a técnica de N-grama de Classe Semântica (NGCS), proposta por Berber Sardinha (2023).

### 2.1 DESIGN, COLETA E LIMPEZA DO CORPUS

O objetivo final deste trabalho é a descrição da linguagem telecinemática norte-americana do ponto-de-vista lexical. Espera-se, ainda, que essa descrição possa vir a informar a produção de materiais didáticos para ensino de inglês como língua estrangeira. A escolha do corpus não teve como objetivo representar a linguagem telecinemática norte-americana, mas sim obter uma amostra dessa linguagem que possa ter utilidade na sala-de-aula de inglês. Assim, foram escolhidos 300 títulos no total, sendo 150 filmes e 150 episódios de séries televisas, em um recorte temporal de 1991 a 2021. Ao final, o corpus totalizou 2.038.377 palavras.

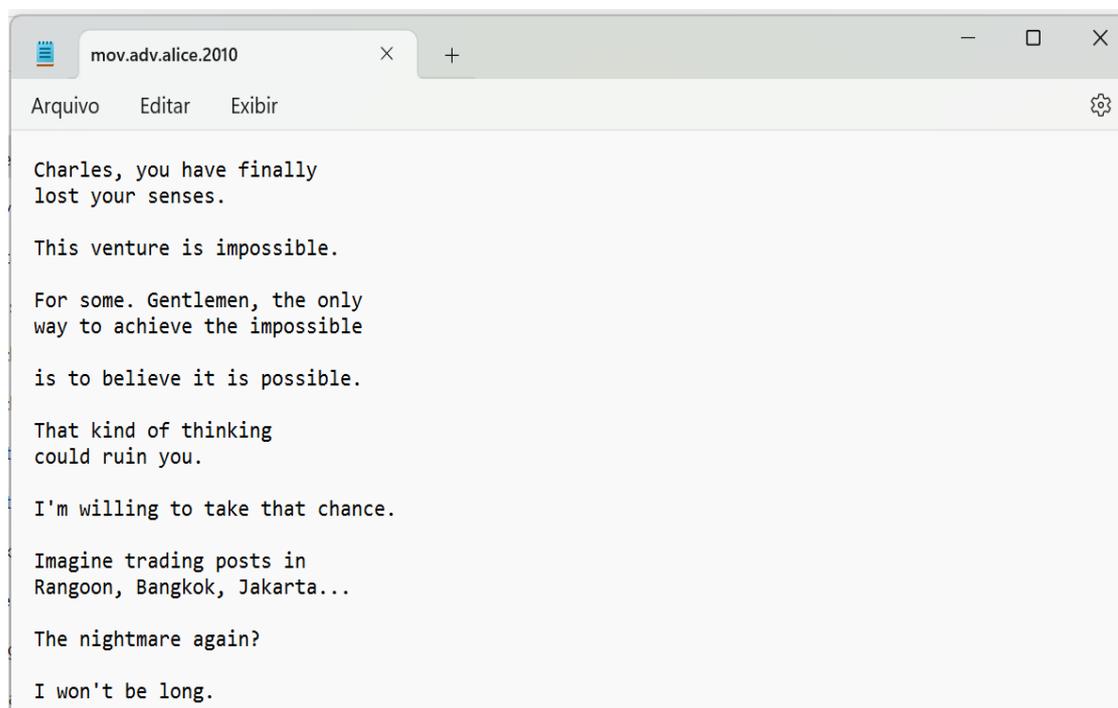
A seleção dos títulos foi feita com base em categorias temáticas, não necessariamente relacionadas a gêneros específicos, tanto para filmes quanto séries de TV. Essa abordagem foi escolhida devido à falta de consenso sobre a composição dos gêneros de TV e cinema, com definições inconsistentes até mesmo dentro da indústria audiovisual (Berber Sardinha, Veirano Pinto, 2020, p. 128). Por exemplo, a série "Downton Abbey" foi premiada como "Melhor Minissérie ou Filme" no 63º Primetime Emmy Awards e como "Melhor Série de Drama" no 2011 BAFTA Awards. Tendo isso em mente, os filmes foram agrupados em cinco categorias: Aventura, Animação, Comédia, Drama e Comédia Romântica. Já para as séries de TV, as cinco categorias foram: Comédia, Investigação Criminal, Drama, Médica e Adolescentes.

A seleção dos títulos dentro de cada categoria foi feita com base nos títulos mais comuns em pesquisas no Google, levando em consideração a popularidade na data da coleta dos dados (dezembro de 2021) e o potencial de engajamento e motivação dos alunos. Foram escolhidos 30 filmes em cada categoria e 2 episódios aleatórios de cada série de TV selecionada.

Após a seleção dos títulos, foi realizada a coleta do corpus de forma manual no site "Subtitles Like Script"<sup>3</sup>, e os arquivos foram salvos em formato .txt. Em seguida, ocorreu a etapa de limpeza dos dados, na qual foram mantidas apenas as falas dos personagens, excluindo-se partes não faladas, como momentos musicais.

Com esse corpus selecionado e limpo, a pesquisa seguiu para as próximas etapas de análise utilizando as técnicas de AMD e NGCS para identificar os padrões linguísticos e discursivos presentes nos discursos telecinemáticos. A partir dessas análises, serão formuladas atividades que visam engajar os alunos na utilização da língua inglesa de forma pragmática, autêntica e funcional, com o propósito de tornar o ensino mais eficiente e motivador.

*Figura 3*



Fonte: Ribeiro (2023) para esta pesquisa.

<sup>3</sup> Do website: <https://subslikescript.com/>

## 2.2 PROCESSAMENTO DO *CORPUS*

O corpus foi segmentado em trechos que correspondiam, em média, a cinco minutos de duração. Esta duração foi escolhida para que sejam facilmente encaixadas dentro de uma aula/hora como parte de uma atividade. Com base em Veirano Pinto (2014), foi determinado que cada minuto de filme norte-americano corresponde a cerca de 79 palavras. Assim, cinco minutos correspondem a 395 palavras, que foram arredondados para 400. Com base nisso, cada texto do corpus foi segmentado em trechos de 400 palavras de extensão. Assim, o corpus compreende 3988 textos segmentados.

Para identificar os NGCSs, o etiquetador semântico USAS foi empregado. Esse etiquetador foi desenvolvido pela Universidade de Lancaster (Rayson, 2007). Um script desenvolvido pelo professor orientador foi utilizado para a etiquetagem automática do corpus.

Após a etiquetagem do corpus, procedeu-se à seleção dos n-gramas de classe semântica mais frequentes no corpus. Nesse processo, o número de ocorrências de cada n-grama foi contabilizado para cada texto.

Em seguida, foram computados os NGCS-chave de cada gênero. Um NGCS-chave é aquele que ocorre mais vezes do que esperado estatisticamente em um gênero do corpus do que nos gêneros restantes do corpus. A abordagem seguida aqui foi inspirada no procedimento de palavras-chave (KeyWords) do pacote de análise de corpus WordSmith Tools (Scott, 1998; Bondi; Scott, 2010; Berber Sardinha, 1999; 2009). No entanto, em vez de considerar a contagem total dos NGCS, foi considerada a contagem dos textos em que os NGCSs ocorriam, seguindo a abordagem de Egbert e Biber (2019). Ao considerar a contagem de textos, evita-se idiosincrasias, isto é, casos em que um mesmo NGCS ocorra muitas vezes no corpus devido a ser usado repetidas vezes em poucos textos.

Para tanto, o corpus foi dividido em subcorpus-alvo e subcorpus de referência. O subcorpus alvo consiste em todos os textos segmentados de cada dos dez gêneros do corpus. Os códigos abaixo são uma representação abreviada que descreve a modalidade e o gênero deste subcorpus usados para categorizar e identificar o conteúdo das obras com base em sua

modalidade (filme ou série de TV, ) e gênero (por exemplo, animação, comédia, drama, etc.), fornecendo informações concisas para sua classificação.

1. m:mov|g:adv: A primeira parte, "m:mov", indica que a modalidade do texto é um filme (em inglês, movie). A segunda parte, "g:adv", revela que o gênero do filme é aventura (adventure).
2. m:mov|g:ani: Representa uma obra na modalidade filme (movie) do gênero animação (animated).
3. m:mov|g:com: Indica uma obra na modalidade filme (movie) do gênero comédia (comedy)
4. m:mov|g:dra: Descreve uma obra na modalidade filme (movie) do gênero drama (drama).
5. m:mov|g:romcom: Refere-se a uma obra na modalidade filme (movie) do gênero comédia romântica (romantic comedy)
6. m:tvsv|g:com: Representa uma obra na modalidade série de TV (TV show) do gênero comédia (comedy)
7. m:tvsv|g:crim: Indica uma obra na modalidade série de TV (TV show) do gênero crime (criminal).
8. m:tvsv|g:dra: Descreve uma obra na modalidade série de TV (TV show) do gênero drama (drama).
9. m:tvsv|g:med: Refere-se a uma obra na modalidade série de TV (TV show) do gênero médico (medical).
10. m:tvsv|g:teen: Representa uma obra na modalidade série de TV (TV show) do gênero adolescente (teen).

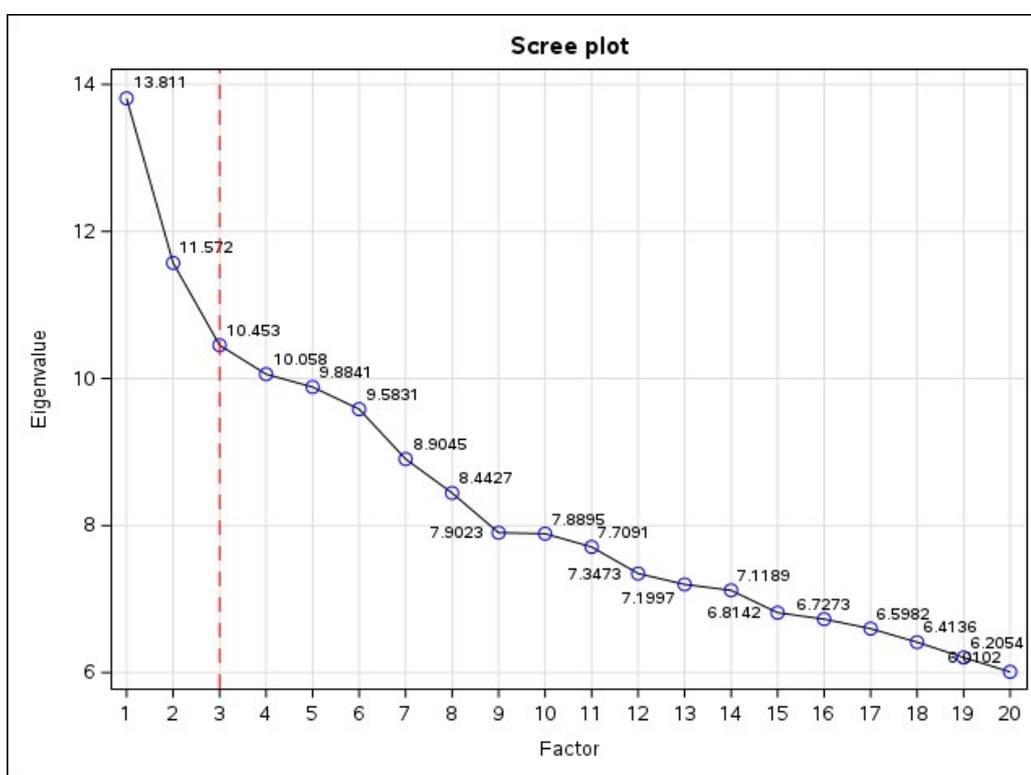
O subcorpus de referência, por sua vez, consiste em todos os textos dos demais gêneros do corpus, com exceção do gênero corresponde ao corpus-alvo. Esse processo foi repetido dez vezes, uma vez para cada gênero. Desse modo, havia dez corpus-alvo e 10 corpus de referência, que se modificavam de acordo com a seleção do corpus-alvo.

Um NGCS foi considerado chave quando sua frequência no corpus alvo (contagem de textos em que ocorre), em comparação com o corpus de referência, atingia um valor de Log-Likelihood igual ao maior que 3,84.

Uma vez obtidos os NGCSs-chave, suas contagens foram coletadas e transpostas para o SAS OnDemand, em que foram realizados os procedimentos estatísticos. Para tanto, foi empregado um programa SAS (script) desenvolvido pelo professor orientador.

Primeiramente, foi realizada análise fatorial sem rotação, a partir da qual foi gerado o *scree plot*, um gráfico que mostra o valor da variação capturada em cada fator. Após a inspeção desse gráfico, optou-se por analisar três fatores, uma vez que a variação nos valores diminui consideravelmente para os próximos fatores (vide figura 4).

Figura 4



Fonte: Ribeiro (2023) para esta pesquisa.

Em seguida, foi realizada uma análise fatorial rotacionada pelo método Promax, por meio da qual foram extraídos três fatores que serão analisados e discutidos.

## CAPÍTULO 3 - ANÁLISE E RESULTADOS

*“People call these things imperfections, but they’re not, that’s the good stuff.” (Good Will Hunting)*

Neste capítulo, serão analisados os fatores encontrados no corpus pela análise estatística. Após a análise qualitativa dos fatores e a inspeção dos textos mais representativos em cada um deles, esses fatores se tornam dimensões, podendo apresentar polos complementares. Em outras palavras, ocorre uma tendência de certas ocorrências serem muito frequentes em um polo e pouco frequentes no outro, e vice-versa. Na AMD, um polo é chamado de positivo e o outro de negativo, porém, é importante destacar que essa nomenclatura é puramente estatística, não envolvendo juízo de valor. Na segunda sessão, será apresentada uma proposta de uso de NGCS em sala de aula.

### 3.1 ANÁLISE DOS DADOS

Nesta seção, descreveremos o discurso subjacente às dimensões encontradas, fornecendo exemplos para cada uma delas. Com base nessa análise, poderemos responder à primeira pergunta da pesquisa:

1. Quais são as dimensões dos n-gramas de classe semânticas presentes no discurso telecinemático?

Para representar os NGCSs que carregaram nos fatores, utilizaremos formulações como "Z5\_G1\_Z5". Essas formulações correspondem à sequência de categorias semânticas identificadas pelo etiquetador USAS. Por exemplo, "Z5\_G1\_Z5" significa "Grammatical bin (Z5): Elections (G1): Grammatical bin (Z5)". Nesse caso, temos um agrupamento de três palavras, duas delas gramatical (Z5), seguida de uma palavra referente ao campo semântico de eleições (G1), seguida de outra palavra gramatical (Z5).

**DIMENSÃO 1 - DISCURSO VOLTADO A INSTITUIÇÕES DO GOVERNO  
VERSUS EXPRESSÃO DE DESEJOS, VONTADES E RELAÇÕES FAMILIARES**

A primeira dimensão revela a presença de dois polos complementares: o polo positivo com 20 variáveis lexicais, e o polo negativo com 13 variáveis. Os NGCSs que carregaram em cada polo aparecem na Tabela 1, que corresponde ao padrão fatorial da dimensão 1. É preciso notar que nessa tabela, bem como nas demais referentes aos padrões fatoriais, os NGCSs entre colchetes tiveram um peso maior em outro fator, não tendo sido usados para computar os escores de fator em questão.

Tabela 1: Padrão fatorial da Dimensão 1

<b>Polo</b>	<b>NGCS</b>
<b>Positivo</b>	Grammatical bin (Z5) : Groups (S5) : Geographical (Z2). (1.05) Elections (G1) : Grammatical bin (Z5) : Grammatical bin (Z5) (0.73) Grammatical bin (Z5) : Elections (G1) : Grammatical bin (Z5) (0.71) [Grammatical bin (Z5) : Personal (Z1) : Residence (H4) (0.70)] Grammatical bin (Z5) : Army (G3) : Direction (M6) (0.55) Grammatical bin (Z5) : Grammatical bin (Z5) : Elections (G1) (0.52) Direction (M6) : Grammatical bin (Z5) : Othernames (Z3) (0.52) Business (I2) : Grammatical bin (Z5) : Grammatical bin (Z5) (0.47) Direction (M6) : Grammatical bin (Z5) : Elections (G1) (0.47) Grammatical bin (Z5) : Grammatical bin (Z5) : Groups (S5) (0.43)

	<p>[Pronouns (Z8) : Army (G3) : Pronouns (Z8) (0.41)]</p> <p>Groups (S5) : Grammatical bin (Z5) : Grammatical bin (Z5) (0.38)</p> <p>Grammatical bin (Z5) : Crime (G2) : Direction (M6)</p> <p>Grammatical bin (Z5) : Business (I2) : Grammatical bin (Z5) (0.38)</p> <p>Grammatical bin (Z5) : Grammatical bin (Z5) : Army (G3) (0.37)</p> <p>Being (A3) : Speech acts (Q2) : Direction (M6) (0.35)</p> <p>Crime (G2) : Grammatical bin (Z5) : Grammatical bin (Z5) (0.35)</p> <p>Direction (M6) : Grammatical bin (Z5) : Groups (S5) (0.35)</p> <p>Groups (S5) : Grammatical bin (Z5) : Pronouns (Z8) (0.32)</p> <p>[Direction (M6) : Pronouns (Z8) : Linear (N4) (0.32)]</p>
<b>Negativo</b>	<p>Grammatical bin (Z5) : Kin (S4) : Kin (S4) (-0.87)</p> <p>[Medical (B3) : Direction (M6) : Grammatical bin (Z5) (-0.66)]</p> <p>Choosing (X7) : Grammatical bin (Z5) : Kin (S4) (-0.62)</p> <p>Pronouns (Z8) : Grammatical bin (Z5) : Drinks (F2) (-0.44)</p> <p>[Medical (B3) : Grammatical bin (Z5) : Pronouns (Z8) (-0.43)]</p> <p>Power (S7) : Pronouns (Z8) : Exclusivizers (A14) (-0.42)</p> <p>Grammatical bin (Z5) : Medical (B3) : Direction (M6) (-0.40)</p> <p>Mental processes (X2) : Pronouns (Z8) : Kin (S4) (-0.40)</p> <p>[Time (T1) : Grammatical bin (Z5) : Liking (E2) (-0.38)]</p>

	<p>Exclusivizers (A14) : Choosing (X7) : Pronouns (Z8) (-0.34)</p> <p>[Grammatical bin (Z5) : Linear (N4) : Speech acts (Q2) (-0.33)]</p> <p>[Sensory (X3) : Pronouns (Z8) : Linear (N4) (-0.31)]</p> <p>[Closed (A10) : Quantities (N5) : Grammatical bin (Z5) (-0.31)]</p>
--	--

Fonte: Ribeiro (2023) para esta pesquisa.

No polo positivo, temos campos semânticos que se referem à interação e dinâmica entre instituições governamentais. As categorias semânticas refletem temas relacionados à política, governo, leis e instituições governamentais. Por outro lado, no polo negativo, encontramos campos semânticos que destacam a forma como as pessoas se relacionam entre si, expressando sentimentos e opiniões pessoais. Os campos semânticos refletem temas emocionais e pessoais, referindo-se a interações interpessoais e sentimentos das personagens.

#### POLO POSITIVO - DISCURSO VOLTADO A INSTITUIÇÕES DO GOVERNO

Ao analisar o polo positivo, foi possível identificar subcategorias distintas relacionadas às diversas formas de se referir à autoridade do governo. A primeira subcategoria diz respeito ao próprio país e suas instituições governamentais, conforme exemplificado no exemplo 1 abaixo. Esse exemplo, retirado do filme *The Post* (2017), traz diversas sequências lexicais relacionadas ao aparato de governo dos EUA.

### Exemplo 1

- Mr. Bradlee, I have **been advised by** the Secretary of Defense that the material published in The Washington Post this morning contains information relating to the national defense of **the United States** and bears a top-secret classification.

- As such, the publication of this information is directly prohibited by the Espionage Act, title 18 of **the United States** code, section 793. As publication will cause irreparable injury... to the defense interests of **the United States**. I respectfully request that you publish no further information of this character and advise me that you have made arrangements for the return of these documents to the Department of Defense.

O exemplo 2, de modo semelhante, também traz sequências lexicais ligadas as guerras das quais os EUA participaram.

- I faked my own death to get out of fighting in World War II. Of course, I went **to prison for** a stretch.

- Wow, you really are a piece of...

- But when I was in the can, I doctored my own conscription papers. Long story short, I spent most **of the war in an Allied** prison.

*Grace and Frankie* (Temporada 6 Episódio 10)

O exemplo 3 traz sequências lexicais referentes a institutos jurídicos.

- What kind of nonsense is this? Put me **on the council and** not make me a master? It's never been done in the history of the Jedi. It's insulting.

- Calm down, Anakin. You have been given a great honor. To be **on the council** at your age, it's never happened before. The fact of the matter is you are too close to the chancellor. The council doesn't like it when he interferes in Jedi affairs.

*Star Wars (2005 )*

Ainda nessa dimensão, observamos NGCSs referentes ao governo e governantes, mesmo de outros países. Por exemplo, no exemplo 4, temos menção ao ministro britânico das finanças, além do presidente norte-americano.

Exemplo 4

- And since bullies only respond to strength, from now onward, I will be prepared to be much stronger. **And the President** should be prepared for that. Mr. President! Joe.

- It's your sister on line four. All right. Yes, I'm very busy and important.

- How can I help you?

- Have you gone completely insane? You can't be sensible all the time.

- You can if you're prime minister.

- Oh, dear, it's **the Chancellor of the Exchequer** on the other line.

- No, it isn't! I'll call you back.

*Love Actually (2003)*

No exemplo 5, há sequências lexicais referentes a instituições históricas que não existem mais.

- Let me be blunt. Will the Southern states resume their former position **in the Union** speedily enough to enable us to block ratification to this 13th Amendment? I'd like peace immediately.

- Yes, and?

- I'd like your states restored to their practical relations with the Union immediately. If this could be given to me in writing, as Vice **President of the Confederacy**, I'd bring that document with celerity, to Jefferson Davis.

*Lincoln* (2012)

Ainda nesse polo, encontramos as formas pelas quais o governo exerce poder sobre os cidadãos, incluindo as forças armadas e as burocracias estabelecidas para a manutenção da ordem.

Exemplo 6

- Two weeks later, I left Vietnam. The ceremony was kicked off with a candid **speech by the President** regarding the need for further escalation **of the war in** Vietnam. President Johnson awarded four Medals of Honor to men from each **of the Armed Services**.

*Forrest Gump* (1994)

Exemplo 7

- He likely is not a college graduate, but feels nonetheless superior to those with advanced education, and he is likely employed **in a bureaucratic** entity, possibly civil service

or quasi - public service, from which he feels alienated. He has a problem **with authority and a** deep-seated resentment of those who he feels have impeded his progress professionally.

*The Wire* (Temporada 5 Episódio 8)

#### Exemplo 8

- We 're on every network **in the country** . What's the protocol? The protocol? I don't know, Bill. What is **the protocol for** an armed celebrity fugitive being chased by 12 police cars and seven helicopters?

*The People v OJ Simpson* (Temporada 1 Episódio 2)

#### Exemplo 9

- You'll not be a conquered people, Mr Hunter. You will be citizens. Returned to the **laws and the** guarantees of rights **of the Constitution**. Which now extinguishes slavery. And with it, our economy. All our laws will be determined **by a Congress** of vengeful Yankees. All our rights will be subject to a Supreme Court benched by bloody Republican radicals.

*Lincoln* (2012)

POLO NEGATIVO - EXPRESSÃO DE DESEJOS, VONTADES E RELAÇÕES FAMILIARES

Diferentemente do polo positivo que descrevia a relação impessoal dentro das instituições governamentais, o polo negativo mostra a demonstração de desejos e vontades. Na primeira subcategoria do polo negativo, encontramos exemplos de expressões de desejos

relacionados a casamento ou a não casamento, não necessariamente implicando a existência de uma segunda pessoa com quem casar.

#### Exemplo 10

- It's just that you **wanted to get married** at The Boathouse in Mom's dress too.
- I know. No, I just - I didn't know that that's what you wanted... but if it is, then I **just want you** to be happy. You know that.

*27 Dresses* (2008)

#### Exemplo 11

- I don't **want to marry** you.
- But you **want to marry** that stranger downstairs?
- No, I don't **want to marry** anyone! Oh, Steve, you're so sweet, and handsome, and surprisingly good at mini golf, and I appreciate you being my first boyfriend, but... Oh, no. Did I just talk you into breaking up with me?
- Sorry, but I think you did.

*Never Have I Ever* (Temporada 1 Episódio 9)

A segunda subcategoria da dimensão 1, ainda dentro do polo negativo, mostra a expressão de desejos e vontades de forma geral. Como podemos ver nos exemplos abaixo, há o frequente uso dos verbos *want* e *wish*, além das conjugações do futuro, como *will* e *going to*.

## Exemplo 12

- I know you're **gonna enjoy** this. I'm gonna have to try and enjoy it even more.
- Just put it down. Take it easy. Take it easy. Drop the weapon now!
- It's my own damn fault, just shoot him!
- Let him go now! Drop it!
- What? Sorry? What do you want?
- I **just want my** phone call. What... ? What...?

*The Dark Knight* (2008)

## Exemplo 13

- I **just wish he** could have been here to **see it finally** happen.
- He saw it, Clark, believe me.

*Man of Steel* (2013)

## Exemplo 14

- And I **just want you** to know, from the bottom of my heart, from this moment forward, I solemnly promise that I will never... Emmet!
- Hey, everyone! Is everyone okay? Where's Lucy?
- Emmet!
- Lucy! We did it.

- Emmet, wait. Batman, there's something I need to say to you. No.
- Wyldstyle... I mean... Lucy. He's the hero you deserve
- Thanks, Batman.

*The Lego Movie* (2014)

#### Exemplo 15

- I **just want you** to **know your father** and I are totally supportive.

*Easy A* (2010)

#### Exemplo 16

- There's got to be something I can do to cheer you up.
- I **just wish I** could see my dog.
- So I 'd just bring the dog in for a quick visit.
- It would mean the world to her.

*Scrubs* (Temporada 6 Episódio 14)

## **DIMENSÃO 2 – DISCURSO VOLTADO AO TRATAMENTO MÉDICO**

Os NGCSs que carregaram nessa dimensão aparecem na Tabela 2, que corresponde ao padrão fatorial da dimensão 2. Conforme observado, os NGCSs entre colchetes tiveram um peso maior em outro fator, não tendo sido usados para computar os escores de fator em questão.

Essa segunda dimensão apresenta um polo interpretável apenas, o negativo, composto por quatorze variáveis, refletindo discurso voltado a instituições hospitalares e ao atendimento médico. O polo positivo, por sua vez, é composto por oito variáveis, três das quais carregaram com peso maior em algum outro fator. Durante a interpretação dos textos, notou-se que as variáveis restantes não tornavam o polo interpretável. Assim, devido a ter poucas variáveis com peso primário no fator em questão interpretáveis no seu conjunto, esse polo foi descartado.

Tabela 2: Padrão fatorial da Dimensão 2

<b>Polo</b>	<b>NGCS</b>
<b>Positivo</b>	Grammatical bin (Z5) : Linear (N4) : Speech acts (Q2) (1.49) Grammatical bin (Z5) : Physical (O4) : Speech acts (Q2) (0.97) Grammatical bin (Z5) : Personal (Z1) : Residence (H4) (0.94) [Grammatical bin (Z5) : Kin (S4) : Kin (S4) (0.68)] [Choosing (X7) : Grammatical bin (Z5) : Kin (S4) (0.57)] [Pulling (M2) : Pronouns (Z8) : Army (G3) (0.41)] Sensory (X3) : Pronouns (Z8) : Linear (N4) (0.32) Speech acts (Q2) : Pronouns (Z8) : Choosing (X7) (0.31)
<b>Negativo</b>	Grammatical bin (Z5) : Medical (B3) : Grammatical bin (Z5) (-0.60) Medical (B3) : Grammatical bin (Z5) : Pronouns (Z8) (-0.53) Disease (B2) : Grammatical bin (Z5) : Pronouns (Z8) (-0.46)

Closed (A10) : Quantities (N5) : Grammatical bin (Z5) (-0.45)
Pronouns (Z8) : Pulling (M2) : Relationship (S3) (-0.41)
[Pronouns (Z8) : Grammatical bin (Z5) : Drinks (F2) (-0.41)]
Grammatical bin (Z5) : Being (A3) : Ability (X9) (-0.41)
Grammatical bin (Z5) : Relationship (S3) : Being (A3) (-0.39)
Direction (M6) : Pronouns (Z8) : Linear (N4) (-0.39)
Direction (M6) : Grammatical bin (Z5) : Medical (B3) (-0.37)
Discourse (Z4) : Being (A3) : Pronouns (Z8) (-0.35)
Grammatical bin (Z5) : Disease (B2) : Grammatical bin (Z5) (-0.32)
Pronouns (Z8) : Sensory (X3) : Liking (E2) (-0.31)
Grammatical bin (Z5) : Grammatical bin (Z5) : Medical (B3) (-0.31)

Fonte: Ribeiro (2023) para esta pesquisa.

## POLO NEGATIVO - DISCURSO VOLTADO AO TRATAMENTO MÉDICO

Diferentemente do polo positivo que reflete diálogos relacionados às instituições sociais, no polo negativo, por outro lado, foi possível reconhecer a presença de sequências lexicais referentes a instituições hospitalares.

No exemplo 17, aparecem sequências lexicais referentes a atendimento e internação hospitalares.

- Do you remember that day **at the hospital** ? Of course you remember that day **at the hospital**. Well, there was a little mix - up. I saw Peter get pushed onto the tracks, and I saved his life. But when I got **to the hospital**, they wouldn't let me see him, so someone told the doctor that I was his fiancée. Only... It 's not true. I was never engaged to Peter.

- Why didn't you say something?

- Because I didn't know how to tell you. We never even met until that day on the tracks. It's just when we were **in the hospital** room, everything happened so fast, and I couldn't tell you the truth.

*While You Were Sleeping* (1995)

No exemplo 18, encontramos sequência lexical referente a equipamento hospitalar para tratamento.

- But other than that, his vitals are stable. And we're hopeful we can get him **off the ventilator**.

- Yes. I know. And Molly's getting her wound redressed.

- I just came to give you this.

- Oh, uh... Dr. Shackleford said you 'd want it.

*Royal Pains* (Temporada 5 Episódio 11)

No exemplo 19, observamos sequências referentes à medicação.

- The pharmacy asked me **for a prescription**, and seeing as how I was only looking for a visual aid, I didn't have one. So I had to drag our chairwoman here **to the dispensary** where we each had to sign wavers affirming our educational intent .

*New Amsterdam* (Temporada 2 Episódio 15)

No exemplo 20, verificamos tanto referência a medicação quanto a infecção.

- You think you have a good idea, but you're going to shut yourself up?

- It could be caused by **an infection from** bacteria that leaked as a result of the puncture. The treatments are diametrically opposed: antibiotics if it's infection, steroids if it's inflammation.

- If she needs **antibiotics and she** needs steroids instead, it will suppress her immune system, causing **the infection to** flourish and kill her. If she needs steroids and we give her

- antibiotics, then...

- I get it. We 'd better be right.

*The Good Doctor* (Temporada 1 Episódio 13)

### **DIMENSÃO 3 – DISCURSO DA VIOLÊNCIA VERSUS DISCURSO SENTIMENTAL**

A terceira dimensão traz, novamente, dois polos. O polo positivo é composto por oito NGCSs relacionadas à criminalidade. Por outro lado, o polo negativo traz 12 NGCSs relacionados a relações interpessoais. Conforme mencionado anteriormente, os NGCSs

inseridos entre colchetes apresentaram uma maior relevância em relação a outro fator, o que resultou na sua exclusão do cálculo dos escores do fator em questão.

Tabela 3: Padrão fatorial da Dimensão 3

<b>Polo</b>	<b>NGCS</b>
<b>Positivo</b>	<p>Pronouns (Z8) : Army (G3) : Pronouns (Z8) (1.09)</p> <p>Medical (B3) : Direction (M6) : Grammatical bin (Z5) (0.99)</p> <p>Pulling (M2) : Pronouns (Z8) : Army (G3) (0.91)</p> <p>Power (S7) : Coming (M1) : Grammatical bin (Z5) (0.61)</p> <p>[Grammatical bin (Z5) : Physical (O4) : Speech acts (Q2) (0.52)]</p> <p>Pronouns (Z8) : Life (L1) : Pronouns (Z8) (0.37)</p> <p>Creatures (L2) : Direction (M6) : Grammatical bin (Z5) (0.34)</p> <p>Getting (A9) : Grammatical bin (Z5) : Army (G3) (0.32)</p>
<b>Negativo</b>	<p>Time (T1) : Grammatical bin (Z5) : Liking (E2) (-0.47)</p> <p>Mental processes (X2) : Liking (E2) : Grammatical bin (Z5) (-0.42)</p> <p>Business (I2) : Grammatical bin (Z5) : Pronouns (Z8) (-0.38)</p> <p>Pronouns (Z8) : Measurement (N3) : Quantities (N5) (-0.36)</p> <p>Grammatical bin (Z5) : Liking (E2) : Pronouns (Z8) (-0.35)</p> <p>Time (T1) : Time (T1) : Coming (M1) (-0.35)</p> <p>Kin (S4) : Grammatical bin (Z5) : Kin (S4) (-0.34)</p>

	Grammatical bin (Z5) : Money (I1) : Abstract (A1) (-0.34)
	Business (I2) : Pronouns (Z8) : Grammatical bin (Z5) (-0.33)
	Grammatical bin (Z5) : Linear (N4) : Choosing (X7) (-0.33)
	Grammatical bin (Z5) : Liking (E2) : Grammatical bin (Z5) (-0.33)
	Linguistics (Q1) : Grammatical bin (Z5) : Pronouns (Z8) (-0.32)

Fonte: Ribeiro (2023) para esta pesquisa.

## POLO POSITIVO – DISCURSO DA VIOLÊNCIA

O exemplo 21 abaixo traz expressões típicas de filmes de ação, envolvendo uso de armas de fogo e acusações de assassinato.

- You could be anywhere in the world right now. Why did you come back? Well, like you said: "Family is everything."
- And you took mine away from me. **You killed my** brother
- Well, the good news is you're about to see him again.
- **Drop your weapon!**
- Your brother was a coward, crying like a baby. He begged for his life. Now you will do the same.
- Never.
- Vicente Cortez, you're under arrest for murder and attempted murder, and I'll figure

out the rest later.

- Miss Riva, put the gun down. If you pull that trigger, you'll go to jail for the rest of your life. Think about what you're doing.
- **He killed my** brother. He killed the only family I had. I can't let him live.
- And I can't let **you kill him**. Don't make me stop you. I'm asking you for the last time. Put the gun down. You won't regret this. I promise.
- Are you kidding me?
- What? You made me not kill him so that you get to kill him?
- I get to kill him? I just saved your life!
- I have been waiting for six years for this moment and you come at the last second and ruin it?
- I just **took a bullet** for you!
- You **took a bullet** for me? It went right through you and hit me!

*Hot Pursuit* (2015)

No exemplo 22, aparecem formulações relacionadas ao uso de armas de fogo.

- The cops would have found my fingerprints within five blocks of that place, they would have picked me up by now.
- What about the casing? Or casings? How...How many times did **you shoot him** anyway?
- Jesse, did you pick up the casings? Did you pick them up?

*Breaking Bad* (Temporada 4 Episódio 4)

## POLO NEGATIVO – DISCURSO SENTIMENTAL

Enquanto o polo positivo demonstra situações relacionadas à luta armada, o polo negativo traz diferentes formas de relações interpessoais.

Os exemplos 23, 24 e 25 trazem representações de relacionamentos entre enamorados, na família e entre amigos.

### Exemplo 23

- I love you in the morning. I love you. **I'll never leave** you. I love you. I **'ll never leave** you. I love you. **I'll never leave** you. I love your mouth . I love your nose. I love your butt. I love your eyes. I love your belly. I love your ears.

*Ruby Sparks* (2012)

### Exemplo 24

- I've just split up with my boyfriend, actually, so I'm back with my **mom and dad** for a while.
- Sorry.
- No, it's fine. I'm well shot of him. He said I was getting fat.
- I beg your pardon? He said no one's **gonna fancy a** girl with thighs the size of big tree trunks. Not a nice guy, actually, in the end.
- No.

*Love Actually* (2003)

Exemplo 25

- Then you've had ten years.
- Right.
- Well, we 've always been like **brother and sister**.
- It's very bad indeed for, for **brothers and sisters** to have sex.
- Except that we 're not **brother and sister**.
- No, exactly.
- Right, but...
- No, so... I've been waiting half my life for you to wake up **and love me**. Having loved you for half a lifetime, I realized when you left, that I had made a bad choice doing that. And now, it's got even trickier because... if... when you were playing in pubs, we were the perfect match. But now... I - I...I'm an actual schoolteacher in actual lowestoft, and you're the world's greatest singer-songwriter
- No, I'm not. –
- Except that you probably are.

*Yesterday* (2019)

### **3.2 DAS DIMENSÕES PARA INFORMAR O ENSINO DE LÍNGUA ESTRANGEIRA**

*“Always pass on what you have learned.” (Star Wars)*

Nesta seção, serão apresentadas sugestões de uso das dimensões para informar o ensino de inglês como língua estrangeira em vários contextos, como institutos de idioma e aulas particulares. Em particular, as sugestões apresentadas tratarão do ensino de vocabulário de língua inglesa por meio de filmes e séries de televisão.

Em primeiro lugar, uma abordagem multidimensional, como a seguida aqui, fornece uma visão abrangente da língua em uso, de tal modo que cada polo de cada dimensão pode ser visto como uma unidade organizadora da linguagem com vistas ao ensino.

Nesta pesquisa, obtivemos cinco polos dimensionais:

Dim. 1 pos.: Discurso voltado a instituições do governo

Dim. 1 neg.: Expressão de desejos, vontades e relações familiares

Dim. 2 neg.: Discurso voltado ao tratamento médico

Dim. 3 pos.: Discurso da violência

Dim. 3 neg.: Discurso sentimental

Assim, o professor pode encarar cada um desses polos como um conjunto integrado de itens de vocabulário a ser trabalhado com os alunos. Por conjunto integrado, queremos dizer que cada polo representa o vocabulário que coocorre nos filmes e programas de televisão. Desse modo, ao apresentar um filme ou show de TV para os alunos, os itens de vocabulário ocorrerão juntos, ou seja, os alunos encontrarão os itens sendo usados na trama a que assistem.

Em segundo lugar, cada polo representa um tipo de linguagem, ou discurso, em que os filmes e séries se baseiam para elaborar a narrativa. Assim, o professor pode centrar os materiais de ensino de vocabulário em torno desses cinco discursos principais: vocabulário em torno de instituições e questões governamentais; vocabulário em torno de desejos e vontades;

vocabulário em torno de tratamento médico; vocabulário em torno da violência; e vocabulário em torno de sentimentos.

Em terceiro lugar, cada NGCS pode ser usado com um conjunto de itens relacionados de vocabulário a ensinar. Por exemplo, o NGCS 'Grammatical bin (Z5) : Kin (S4) : Kin (S4)' que carregou no polo 'Dim. 1 neg.: Expressão de desejos, vontades e relações familiares', materializa-se desta forma nos filmes e séries:

and get married

both getting married

into getting married

of getting married

a foster parent

the foster parents

to get married

with Mama Mother

with Mummy Mother

with divorced parents

Com base nesse conjunto, o professor pode desenvolver atividades de ensino das características lexicais dos itens. Como esses itens aparecem nos filmes e séries, os alunos poderão verificar como são usados (além de sua pronúncia e entonação) nesses materiais.

Por fim, o professor pode utilizar os NGCSs como ponto de partida para explorar os padrões lexicais, usando outro corpus, como o COCA. Por exemplo, os alunos podem explorar

os padrões da palavra 'parents' no COCA, gerando concordâncias. Para ilustrar, as ocorrências mais frequentes de NOUN + parents (i.e. a palavra parents antecedida de um substantivo) são:

BIRTH PARENTS	609
SCHOOL PARENTS	86
TEEN PARENTS	81
HELICOPTER PARENTS	76
SURROGATE PARENTS	69
CLASS PARENTS	53
CITIZEN PARENTS	49
MINORITY PARENTS	45
LATINO PARENTS	43

Os alunos descobrirão, por exemplo, que 'birth parents' é a forma mais frequente do padrão NOUN + parents. Podem descobrir também que 'class parents' é mais frequentemente 'middle class parents'.

Enfim, cada um dos itens de cada um dos NGCSs pode ser usado como uma oportunidade para aprendizado de vocabulário por meio de corpus.

## CONSIDERAÇÕES FINAIS

*“Make it simple, but significant.” (Mad Men)*

O objetivo desta pesquisa foi investigar e compreender a linguagem telecinemática, explorando suas dimensões através da análise de N-Gramas de Classes Semânticas (NGCS), com o intuito de contribuir para o enriquecimento do Portal das Humanidades. Para alcançar esse objetivo, foram estabelecidos três objetivos específicos e formuladas duas perguntas de pesquisa: 1. Quais as dimensões de N-Gramas de Classes Semânticas presentes no discurso telecinemático? 2. Como as dimensões identificadas podem informar o ensino de vocabulário de inglês como língua estrangeira?

A fim de responder essas perguntas, foi desenhado e coletado um corpus de 300 textos, incluindo 150 filmes e 75 séries, com 2 episódios de cada série, por meio buscas na web. Esse corpus possibilitou a análise e identificação das dimensões que compõem a linguagem telecinemática.

Foram identificadas três dimensões:

Dim. 1: Discurso voltado a instituições do governo versus expressão de desejos, vontades e relações familiares.

Dim. 2: Discurso voltado ao tratamento médico.

Dim. 3: Discurso da violência versus discurso sentimental.

A análise qualitativa detalhada dos NGCS possibilitou o entendimento de como a linguagem telecinemática reflete e representa questões sociais, culturais e emocionais .

Na primeira dimensão, o polo positivo abrange aspectos políticos, institucionais e históricos, fornecendo insights sobre o aparato governamental dos EUA, guerras passadas e figuras de destaque. Além disso, essa dimensão tratou do exercício do poder governamental sobre os cidadãos, incluindo referências às forças armadas e ao aparato burocrático. O polo negativo dessa dimensão, destaca o retrato mais pessoal e subjetivo dos desejos e aspirações dos personagens, acrescentando profundidade emocional à narrativa.

A segunda dimensão analisada revelou dois polos distintos, o positivo e o negativo, relacionados ao discurso sobre instituições sociais e hospitais, respectivamente. O polo positivo, embora inicialmente composto por oito variáveis, foi descartado devido à falta de interpretabilidade de suas variáveis restantes. Já o polo negativo mostrou claramente a presença de sequências lexicais relacionadas a instituições hospitalares, como atendimento médico, internação, equipamento para tratamento e referências a medicações e infecções. Esses resultados destacam a importância dessas sequências lexicais específicas na construção do discurso sobre o ambiente hospitalar, enriquecendo a compreensão dos textos analisados.

A terceira dimensão apresenta dois polos distintos. O polo positivo é composto por oito NGCSs relacionadas à criminalidade, mas algumas delas foram excluídas por serem mais relevantes em outro fator. Em contraste, o polo negativo abrange 12 NGCSs ligadas a relações interpessoais. Os exemplos no polo positivo retratam situações típicas de filmes de ação, envolvendo uso de armas de fogo e acusações de assassinato. Por outro lado, os exemplos no polo negativo mostram diferentes formas de relações afetivas entre enamorados, na família e entre amigos.

A presença frequente do campo semântico (Z5) nos exemplos analisados pode ser atribuída ao caráter oralizado dos textos utilizados, visto que foram escritos de forma a simular a fala espontânea dos personagens nas produções telecinemáticas. Esse aspecto é comum em filmes e séries de TV, onde a linguagem utilizada muitas vezes busca refletir a forma como as pessoas realmente falam no dia a dia (Berber Sardinha; Veirano Pinto, 2019; Veirano Pinto, 2014).

Além disso, é interessante notar que, apesar de termos escolhido filmes e séries de TV como fontes distintas para a análise, não encontramos diferenças significativas nos discursos presentes em ambas as mídias. Os exemplos das três dimensões analisadas mostraram que os discursos são tão semelhantes que encontramos ocorrências relevantes em ambos os casos. Essa descoberta pode ser explicada pela natureza da linguagem telecinemática, que muitas vezes segue padrões e convenções de linguagem verbal semelhantes para a cinematografia e televisão. Os roteiristas, ao criarem diálogos para filmes e séries, podem se basear em estruturas linguísticas e temáticas comuns que são frequentemente utilizadas em ambos os meios.

Essa falta de diferença significativa nos discursos entre filmes e séries é uma constatação importante, pois significa que as atividades e estratégias de ensino propostas com base nas dimensões identificadas podem ser igualmente aplicadas em ambas as mídias. Isso proporciona aos educadores mais flexibilidade e possibilidades ao incorporar a linguagem telecinemática no ensino de inglês como língua estrangeira, permitindo que os alunos explorem uma variedade de conteúdos e temas através de filmes e séries de TV de forma igualmente eficaz.

De fato, um ponto importante a ser destacado é a repetição das relações entre pessoas e diferentes instituições sociais nas dimensões analisadas. Embora essas relações apareçam em mais de uma dimensão, é notável que as formas de interação variam significativamente, demonstrando a diversidade presentes no discurso telecinemático.

Nos polos positivos da primeira e terceira dimensão, encontramos relações interpessoais. Entretanto, essas relações são abordadas de maneiras distintas em cada dimensão. Na primeira dimensão, elas são retratadas por meio de expressões de desejos e vontades, que podem envolver relacionamentos amorosos, sonhos pessoais ou aspirações individuais. Já na terceira dimensão, as formas de interação entre as pessoas são mais evidentes e estão relacionadas a relações afetivas entre enamorados, na família e entre amigos

As considerações apresentadas acerca de como as dimensões poderiam informar o ensino de vocabulário do inglês tentaram mostrar como uma visão multidimensional da língua em uso pode ser um instrumento útil para o professor ou desenvolvedor de materiais de ensino. O conjunto das dimensões serve como uma síntese do conteúdo principal dos filmes e séries, ao mesmo tempo em que cada polo de cada dimensão fornece os itens a serem enfocados, de tal modo que esses itens são naturalmente encontrados juntos nos filmes e séries que o professor leva para a sala de aula. Além disso, os itens individuais que carregaram nas dimensões são eles mesmos trampolins para os alunos explorarem o vocabulário em torno de temas definidos pelas tramas dos filmes e séries.

Ao explorar tópicos presentes em filmes e séries de TV, tais atividades têm o potencial de despertar o interesse e a motivação dos alunos, tornando a aprendizagem da língua inglesa mais atrativa e significativa. O uso da linguagem telecinemática como base para as atividades permite que os alunos se aproximem da língua de forma autêntica e real, uma vez que esses discursos refletem a comunicação cotidiana e atual.

Ao compartilhar as atividades no Portal das Humanidades, este trabalho busca disseminar o conhecimento acadêmico e proporcionar aos educadores de língua inglesa recursos práticos e eficazes para suas aulas. A disponibilidade dessas atividades para outros professores e pesquisadores permite que mais pessoas se beneficiem dessa abordagem dinâmica e relevante para o ensino de línguas, contribuindo para o aprimoramento do ensino e da pesquisa nas Ciências Humanas.

Além disso, a pesquisa destaca a importância de explorar o discurso telecinemático de forma contextualizada e significativa em sala de aula. O uso de filmes e séries de TV como recursos didáticos permite que os educadores abordem temas relevantes da sociedade contemporânea, ao mesmo tempo em que desenvolvem as habilidades linguísticas dos alunos. Essa abordagem autêntica torna o processo de aprendizagem mais enriquecedor e prazeroso, incentivando os alunos a se envolverem mais ativamente na aquisição da língua inglesa.

A utilização da Linguística de Corpus para informar o ensino de inglês como língua estrangeira é fruto do desejo genuíno desta pesquisadora e professora, que busca constantemente aprimorar suas práticas pedagógicas e motivar seus alunos a se interessarem pela língua e cultura inglesas. A experiência de mais de uma década no ensino da língua inglesa permitiu identificar a necessidade de tornar as aulas mais atrativas e estimulantes para os alunos, e as considerações pedagógicas expostas neste trabalho representam uma resposta a essa demanda.

Por fim, a abordagem inovadora das dimensões de N-Gramas de Classes Semânticas (Berber Sardinha, 2023) contribui para o entendimento da linguagem telecinemática de forma mais aprofundada e estruturada. As dimensões identificadas nas análises fornecem uma visão abrangente dos principais temas e elementos presentes nos discursos telecinemáticos, o que possibilita uma compreensão mais completa da diversidade linguística e cultural retratada nas produções audiovisuais.

## REFERÊNCIAS

ARAÚJO, L. R.; VOSS, R. C. R. *Cinema em sala de aula: Identificação e projeção no ensino/aprendizagem da língua inglesa*. Conexão, Comunicação e Cultura, UCS, Caxias do Sul, v. 8, n. 15, jan./jun. 2009.

ARCHER, D.; WILSON, A.; RAYSON, P. *Introduction to the USAS category system*. Benedict project report, 2002.

ARCHER, D.; RAYSON, P.; PIAO, S.; McENERY, T. *Comparing the UCREL Semantic Annotation Scheme with lexicographical taxonomies*. In G. Williams & S. Vessier (Eds.), *Proceedings of the 11th EURALEX (European Association for Lexicography) International Congress (Euralex 2004) (Vol. III, pp. 817–827)*. Lorient: Université de Bretagne Sud, 2004.

BELLONI, M. L. *A televisão como ferramenta pedagógica na formação de professores*. Educação e Pesquisa, São Paulo, v. 29, n. 2, p. 287-301, jul./dez. 2003.

BERBER SARDINHA, T. *Usando WordSmith Tools na pesquisa Linguística*. DIRECT Papers 40. LAEL, PUC/SP e AELSU, University of Liverpool. 1999. Disponível na Internet: [www.direct.f2s.com](http://www.direct.f2s.com). Acesso em: 28 de julho de 2023.

\_\_\_\_\_. *Análise multidimensional*. In: D.E.L.T.A. n.16, v.1, p. 99-127. 2000.

\_\_\_\_\_. *Linguística de Corpus*. São Paulo: Manole, 2004. 410 p.

\_\_\_\_\_. *Pesquisa em Linguística de Corpus com WordSmith Tools*. Campinas: Mercado de Letras. 2009.

\_\_\_\_\_. *25 years later: Comparing Internet and pre-Internet registers*. In: BERBER SARDINHA, T.; VEIRANO PINTO, M. (Eds.). *Multi-dimensional analysis, 25 years on. A tribute to Douglas Biber*. Studies in Corpus Linguistics, v. 60. Amsterdam/Philadelphia: John Benjamins, 2014. p. 81–105.

\_\_\_\_\_. *Dimensions of variation across Internet registers*. International Journal of Corpus Linguistics, v. 23, n. 2, p. 125–157, 2018.

\_\_\_\_\_. Discourse of academia from a multi-dimensional perspective. In: FRIGINAL, E. e HARDY, Jack. *The Routledge handbook of corpus approaches to discourse analysis*. Abingdon: Routledge, 2021. p. 298-318.

\_\_\_\_\_. *Lexical bundles*. Seminário. UFRGS. Porto Alegre. 2023.

BERBER SARDINHA, T.; MADUREIRA, S.; BRAIT, B.; SOUZA-E-SILVA, M. C. P.; CAMARGO, Z.; SPAGNUOLO, R. L.; BRAZ, A. A. B. *Portal multimodal/multilíngue para o avanço da ciência aberta nas humanidades*. 2021.

BERBER SARDINHA, T.; KAUFFMANN, C.H.; ACUNZO, C. *A multidimensional analysis of register variation in Brazilian Portuguese*. *Corpora*, v.9, n.2, p. 239-271, 2014.

BERBER SARDINHA, T.; SHIMAZUMI, M. *Dimensions of register variation in British television*. In: Asia-Pacific Corpus Linguistics Conference, 2020, Yon-sei University, Seoul, South Korea.

BERBER SARDINHA, T.; VEIRANO PINTO, M. (Eds.). *Multi-dimensional analysis, 25 years on. A tribute to Douglas Biber*. Amsterdam; Philadelphia: John Benjamins, 2014.

\_\_\_\_\_. *Multi-dimensional analysis*. Cambridge: Bloomsbury Academic, 2019.

\_\_\_\_\_. *A linguistic typology of American television*. *International Journal of Corpus Linguistics*, v. 26, n. 1, p. 127-160, 2020.

\_\_\_\_\_. *Dimensions of variation across American television registers*. *International Journal of Corpus Linguistics*, v. 24, n. 1, p. 3-32, 2019. Comunicação.

BIBER, D. *Variation across speech and writing*. Cambridge: Cambridge University Press, 1988.

BIBER, D.; CONRAD, S.; REPPEN, R. *Corpus linguistics - Investigating language structure and use*. Cambridge: Cambridge University Press, 1998.

BONDI, M.; SCOTT, M. (Eds.). *Keyness in Texts*. John Benjamins. 2010.

BROGINI, A. A. *Representações Contemporâneas da Sustentabilidade: Uma Análise Multidimensional Lexical Discursiva como Contribuição para o Portal*

*Multimodal/Multilíngue para o Avanço da Ciência Aberta nas Humidades*. Dissertação de Mestrado - LAEL. PUC. São Paulo, 2023.

BROWN, H. D. *Teaching by principles: An Interactive Approach to Language Pedagogy*. 3. ed. Pearson Education ESL, 2007.

DEHAENE, S. *É assim que aprendemos: Por que o cérebro funciona melhor que qualquer máquina (ainda...)*. Tradução de Rodolfo Ilari. São Paulo: Context, 2022.

DELFINO, M. C. N. *Uso de música para o ensino de inglês como língua estrangeira em um ambiente baseado em corpus*. Dissertação de Mestrado - LAEL. PUC. São Paulo, 2016.

\_\_\_\_\_. *More than words: análise multimodal multidimensional da música popular em língua inglesa*. Tese de Doutorado - LAEL. PUC. São Paulo, 2022.

DEUS, A. I. S. *A linguagem cinematográfica na formação docente: Cinema e educação na ação pedagógica*. In: XI Seminário de Iniciação à Docência, Natal/RS, 2014.

EGBERT, J.; BIBER, D. *Incorporating text dispersion into keyword analyses*. *Corpora*, v. 14, n. 1, p. 77-104, 2019.

FABRIS, E. H. *Cinema e educação: Um caminho metodológico*. Universidade Federal do Rio Grande do Sul. Porto Alegre: Educação e Realidade, 2008.

FERREIRA, A. M. *Cinema e educação: Uma reflexão sobre a formação de educadores na/para a linguagem audiovisual*. Dissertação de Mestrado, Universidade Federal de Juiz de Fora, Juiz de Fora, 2009.

\_\_\_\_\_. *Cinema e Formação de Professores*. *Travessias, UNIOESTE*, v. 6, p. 3, 2009.

FREIRE, P. *Educação como prática da liberdade*. 45. ed. São Paulo: Paz e Terra, 2019.

FRANCO, B. A. R. *O uso de séries em sala de aula como recurso para ensino de língua inglesa*. *Revista CBTecLE*, nº 2, v. 1, n. 2, p. 176-191, 2018.

GIL, B. C. *A incidência do princípio idiomático e do princípio da escolha aberta na produção escrita de alunos brasileiros de inglês como língua estrangeira*. Dissertação de Mestrado, LAEL, PUC, São Paulo, 2014.

KAUFFMANN, C. H. *Linguística de corpus e estilo: análises multidimensional e canônica na ficção de Machado de Assis*. Tese de Doutorado, LAEL, PUCSP, São Paulo, 2020.

KAUFFMANN, C. H.; BERBER SARDINHA, T. *Brazilian Portuguese literary style*. In: FRIGINAL, E.; HARDY, J. (Eds.). *The Routledge handbook of corpus approaches to discourse analysis*. Abingdon: Routledge, 2021. p. 354-375.

HOLLEBEN, I. M. A. D. S. *Cinema e educação: Um diálogo possível*. Ponta Grossa, 2007.

LEFFA, V. J. *Metodologia do ensino de línguas*. In: BOHN, H. I.; VANDRESEN, P. (Eds.). *Tópicos em Linguística aplicada: O Ensino de Línguas Estrangeiras*. Florianópolis: Ed. da UFSC, 1988. p. 211-236.

\_\_\_\_\_. *Produção de materiais de ensino: Teoria e Prática*. 2. ed. Pelotas: Educat, 2008.

MAYER, C. *O que e como escrevemos na web: um estudo multidimensional de variação de registro em língua inglesa*. Tese de Doutorado - LAEL, PUCSP. São Paulo, 2020.

QUINTA JUNIOR, E. R. *Leni Riefenstahl (1935): O cinema documental nazista*. Dissertação de Mestrado - Universidade Federal de Goiás, Faculdade de História (FH), Programa de Pós-Graduação em História, Goiânia, 2019.

RAYSON, P. *Wmatrix: A web-based corpus processing environment*. Computing Department, Lancaster University, 2007. Disponível em: <http://ucrel.lancs.ac.uk/wmatrix/>. Acesso em 28 de julho de 2023.

ROMEIRO Y. M. T. D. *A linguagem verbal das artes visuais: uma análise multidimensional do discurso sobre a fotografia de Sally Mann*. Dissertação de Mestrado - LAEL, PUCSP. São Paulo, 2020.

SCOTT, M. *WordSmith Tools*. In (Version 3) Oxford University Press, 1998.

SETTON, M. G. *Mídia e educação*. 1. ed. São Paulo: Editora Contexto, 2011.

THE NEW LONDON GROUP. *A pedagogy of multiliteracies: Designing Social Futures*. Harvard Educational Review; Spring 1996; p. 60 - 92.

VEIGA, A. T. *As dimensões da fé: sete religiões mundiais em uma análise multidimensional lexical*. Tese de Doutorado - LAEL, PUCSP. São Paulo, 2021.

VEIRANO PINTO, M. *A linguagem dos filmes norte-americanos ao longo dos anos: uma Abordagem Multidimensional*. Tese de Doutorado - LAEL, PUCSP. São Paulo, 2013.

\_\_\_\_\_. *Dimensions of variation in North American movies*. In: BERBER SARDINHA, T. e VEIRANO PINTO, M. (Eds). *Multi-Dimensional Analysis, 25 years on: A Tribute to Douglas Biber*. Amsterdam/Philadelphia, PA: John Benjamins, 2014. p. 109-149.

## ANEXOS

### ANEXO 1: TABELA COMPLETA DE CATEGORIAS E SUBCATEGORIAS SEMÂNTICAS

A1	General and abstract terms
A1.1.1	General actions, making etc.
A1.1.2	Damaging and destroying
A1.2	Suitability
A1.3	Caution
A1.4	Chance, luck
A1.5	Use
A1.5.1	Using
A1.5.2	Usefulness
A1.6	Physical/mental
A1.7	Constraint
A1.8	Inclusion/Exclusion
A1.9	Avoiding
A2	Affect
A2.1	Affect:- Modify, change
A2.2	Affect:- Cause/Connected
A3	Being
A4	Classification
A4.1	Generally kinds, groups, examples
A4.2	Particular/general; detail
A5	Evaluation
A5.1	Evaluation:- Good/bad
A5.2	Evaluation:- True/false
A5.3	Evaluation:- Accuracy
A5.4	Evaluation:- Authenticity
A6	Comparing
A6.1	Comparing:- Similar/different
A6.2	Comparing:- Usual/unusual
A6.3	Comparing:- Variety
A7	Definite (+ modals)
A8	Seem
A9	Getting and giving; possession
A10	Open/closed; Hiding/Hidden; Finding; Showing
A11	Importance
A11.1	Importance: Important
A11.2	Importance: Noticeability
A12	Easy/difficult
A13	Degree
A13.1	Degree: Non-specific
A13.2	Degree: Maximizers
A13.3	Degree: Boosters
A13.4	Degree: Approximators

- A13.5 Degree: Compromisers
- A13.6 Degree: Diminishers
- A13.7 Degree: Minimizers
- A14 Exclusivizers/particularizers
- A15 Safety/Danger
- B1 Anatomy and physiology
- B2 Health and disease
- B3 medicines and medical treatment
- B4 Cleaning and personal care
- B5 Clothes and personal belongings
- C1 Arts and crafts
- E1 Emotional actions, states and processes
- E2 Liking
- E3 Calm/Violent/Angry
- E4 Happy/sad
- E4.1 Happy/sad: Happy
- E4.2 Happy/sad: Contentment
- E5 Fear/bravery/shock
- E6 Worry, concern, confident
- F1 Food
- F2 Drinks
- F3 Cigarettes and drugs
- F4 Farming & Horticulture
- G1 Government, Politics and elections
- G1.1 Government etc.
- G1.2 Politics
- G2 Crime, law and order
- G2.1 Crime, law and order: Law and order
- G2.2 General ethics
- G3 Warfare, defence and the army; weapons
- H1 Architecture and kinds of houses and buildings
- H2 Parts of buildings
- H3 Areas around or near houses
- H4 Residence
- H5 Furniture and household fittings
- I1 Money generally
- I1.1 Money: Affluence
- I1.2 Money: Debts
- I1.3 Money: Price
- I2 Business
- I2.1 Business: Generally
- I2.2 Business: Selling
- I3 Work and employment
- I3.1 Work and employment: Generally
- I3.2 Work and employmeny: Professionalism
- I4 Industry
- K1 Entertainment generally
- K2 Music and related activities
- K3 Recorded sound etc.

K4	Drama, the theatre and showbusiness
K5	Sports and games generally
K5.1	Sports
K5.2	Games
K6	Childrens games and toys
L1	Life and living things
L2	Living creatures generally
L3	Plants
M1	Moving, coming and going
M2	Putting, taking, pulling, pushing, transporting &c.
M3	Vehicles and transport on land
M4	Shipping, swimming etc.
M5	Aircraft and flying
M6	Location and direction
M7	Places
M8	Remaining/stationary
N1	Numbers
N2	Mathematics
N3	Measurement
N3.1	Measurement: General
N3.2	Measurement: Size
N3.3	Measurement: Distance
N3.4	Measurement: Volume
N3.5	Measurement: Weight
N3.6	Measurement: Area
N3.7	Measurement: Length & height
N3.8	Measurement: Speed
N4	Linear order
N5	Quantities
N5.1	Entirety; maximum
N5.2	Exceeding; waste
N6	Frequency etc.
O1	Substances and materials generally
O1.1	Substances and materials generally: Solid
O1.2	Substances and materials generally: Liquid
O1.3	Substances and materials generally: Gas
O2	Objects generally
O3	Electricity and electrical equipment
O4	Physical attributes
O4.1	General appearance and physical properties
O4.2	Judgement of appearance (pretty etc.)
O4.3	Colour and colour patterns
O4.4	Shape
O4.5	Texture
O4.6	Temperature
P1	Education in general
Q1	Linguistics actions, states and processes
Q1.2	Paper documents and writing
Q1.3	Telecommunications

Q2	Speech acts
Q2.1	Speech etc:- Communicative
Q2.2	Speech acts
Q3	Language, speech and grammar
Q4	The Media
Q4.1	The Media:- Books
Q4.2	The Media:- Newspapers etc.
Q4.3	The Media:- TV, Radio and Cinema
S1	Social actions, states, and processes
S1.1	Social actions, states, and processes
S1.1.1	Social actions, states, and processes
S1.1.2	Reciprocity
S1.1.3	Participation
S1.1.4	Deserve etc.
S1.2	Personality traits
S1.2.1	Approachability and Friendliness
S1.2.2	Avarice
S1.2.3	Egoism
S1.2.4	Politeness
S1.2.5	Toughness; strong/weak
S1.2.6	Sensible
S2	People
S2.1	People:- Female
S2.2	People:- Male
S3	Relationship
S3.1	Relationship: General
S3.2	Relationship: Intimate/sexual
S4	Kin
S5	Groups and affiliation
S6	Obligation and necessity
S7	Power relationship
S7.1	Power, organizing
S7.2	Respect
S7.3	Competition
S7.4	Permission
S8	Helping/hindering
S9	Religion and the supernatural
T1	Time
T1.1	Time: General
T1.1.1	Time: General: Past
T1.1.2	Time: General: Present; simultaneous
T1.1.3	Time: General: Future
T1.2	Time: Momentary
T1.3	Time: Period
T2	Time: Beginning and ending
T3	Time: Old, new and young; age
T4	Time: Early/late
W1	The universe
W2	Light

W3	Geographical terms
W4	Weather
W5	Green issues
X1	Psychological actions, states and processes
X2	Mental actions and processes
X2.1	Thought, belief
X2.2	Knowledge
X2.3	Learn
X2.4	Investigate, examine, test, search
X2.5	Understand
X2.6	Expect
X3	Sensory
X3.1	Sensory:- Taste
X3.2	Sensory:- Sound
X3.3	Sensory:- Touch
X3.4	Sensory:- Sight
X3.5	Sensory:- Smell
X4	Mental object
X4.1	Mental object:- Conceptual object
X4.2	Mental object:- Means, method
X5	Attention
X5.1	Attention
X5.2	Interest/boredom/excited/energetic
X6	Deciding
X7	Wanting; planning; choosing
X8	Trying
X9	Ability
X9.1	Ability:- Ability, intelligence
X9.2	Ability:- Success and failure
Y1	Science and technology in general
Y2	Information technology and computing
Z0	Unmatched proper noun
Z1	Personal names
Z2	Geographical names
Z3	Other proper names
Z4	Discourse Bin
Z5	Grammatical bin
Z6	Negative
Z7	If
Z8	Pronouns etc.
Z9	Trash can
Z99	Unmatched

Fonte: Ribeiro (2023) para esta pesquisa, adaptado de Archer et. al., p. 2, 2004.

ANEXO 2: LISTA COMPLETA DAS NGCS

Z5_S5_Z2	G1_Z5_Z5	Z5_G1_Z5	Z5_G3_M6	Z5_Z5_G1	M6_Z5_Z3	I2_Z5_Z5
M6_Z5_G1	Z5_Z5_S5	S5_Z5_Z5	Z5_G2_M6	Z5_I2_Z5	Z5_Z5_G3	A3_Q2_M6
G2_Z5_Z5	M6_Z5_S5	S5_Z5_Z8	Z5_S4_S4	X7_Z5_S4	Z8_Z5_F2	S7_Z8_A14
Z5_B3_M6	X2_Z8_S4	A14_X7_Z8	Z5_B3_Z5	B3_Z5_Z8	B2_Z5_Z8	A10_N5_Z5
Z8_M2_S3	Z5_A3_X9	Z5_S3_A3	M6_Z8_N4	M6_Z5_B3	Z4_A3_Z8	Z5_B2_Z5
Z8_X3_E2	Z5_Z5_B3	Z8_G3_Z8	B3_M6_Z5	M2_Z8_G3	S7_M1_Z5	Z8_L1_Z8
L2_M6_Z5	A9_Z5_G3	T1_Z5_E2	X2_E2_Z5	I2_Z5_Z8	Z8_N3_N5	Z5_E2_Z8
T1_T1_M1	S4_Z5_S4	Z5_I1_A1	I2_Z8_Z5	Z5_N4_X7	Z5_E2_Z5	Q1_Z5_Z8

Fonte: Ribeiro (2023) para esta pesquisa.