

Um panorama sobre o ensino de Linguística Computacional nos cursos de Letras e Linguística no Brasil

An overview of Computational Linguistics teaching in Letters and Linguistics courses in Brazil

Thiago Blanch PIRES*

Leonel Figueiredo de ALENCAR**

RESUMO: A recente proliferação de tecnologias de inteligência artificial generativa indica a necessidade cada vez mais urgente de integrar a linguística computacional na formação de profissionais de Letras. Para isso, é necessário investigar o que está sendo feito a esse respeito no contexto nacional. Assim, o presente estudo investiga o perfil de ensino-aprendizagem de linguística computacional em cursos de graduação e pós-graduação em Letras no Brasil, analisando materiais bibliográficos e planos de ensino. A linguística computacional abrange tanto motivações científicas quanto tecnológicas, sendo uma área interdisciplinar com terminologia variada que reflete ênfases teóricas e práticas distintas. Pelo que pudemos apurar, inexistem publicações sobre o ensino de linguística computacional no Brasil. Estudos internacionais, ainda que escassos, como os de Amaro (2016), Baldridge e Erk (2008), e Fosler-Lussier (2008), revelam desafios relacionados à aversão a representações formais e à heterogeneidade dos conhecimentos dos alunos. O caso da Universidade de Tartu (Koit *et al.*, 2002) demonstra um modelo interdisciplinar bem-sucedido, o que reforça a importância de se investigar o contexto brasileiro para aprimorar a formação de estudantes de Letras. A metodologia da presente investigação é de método misto, baseada na análise de conteúdo e documental (Moraes, 1999; Libâneo, 1994). Foram coletados e sistematizados programas de ensino de linguística computacional de cursos de Letras em universidades brasileiras, disponíveis publicamente. A análise enfocou ementas, objetivos, metodologias e referências bibliográficas, buscando padrões e tendências pedagógicas. Os resultados indicam a presença da linguística computacional em 3 cursos de graduação e 5 na pós-graduação. A sistematização dos dados em planilha permitiu identificar as referências mais recorrentes e analisar o caráter didático-pedagógico dos programas. Este levantamento inicial sugere um reconhecimento da área, mas ressalta a necessidade de aprofundar a compreensão das práticas de ensino. A presença da disciplina em diversas instituições é um indicativo positivo, mas a escassez de literatura específica sobre o assunto e a complexidade dos desafios pedagógicos demandam

* Doutorado em Ciência da Informação (UnB). Professor Associado do Departamento de Línguas Estrangeiras e Tradução e do Programa de Pós-Graduação em Linguística da Universidade de Brasília (UnB). Brasília, DF – Brasil. pirestb@unb.br

** Doutorado em Linguística pela Universität Konstanz, Alemanha. Professor Titular do Departamento de Letras Estrangeiras e do Programa de Pós-Graduação em Linguística da Universidade Federal do Ceará (UFC). Fortaleza, CE – Brasil. leonel.de.alencar@ufc.br

investigações futuras. Dessa forma, este trabalho serve como um ponto de partida para pesquisas que visem aprimorar currículos e a formação de profissionais de Letras.

PALAVRAS-CHAVE: Linguística Computacional. Ensino Superior. Letras. Linguística. Processamento de Linguagem Natural.

ABSTRACT: The recent proliferation of generative artificial intelligence technologies indicates the increasingly urgent need to integrate computational linguistics into the training of language professionals. To achieve this, it is necessary to investigate what is being done in this regard in the national context. Therefore, this study investigates the teaching and learning profile of computational linguistics in undergraduate and graduate Letters and Linguistics programs in Brazil, analyzing bibliographical materials and syllabi. Computational linguistics encompasses both scientific and technological motivations, being an interdisciplinary field with varied terminology that reflects theoretical and practical emphases. The literature on the subject points to a gap in the national literature on the teaching of computational linguistics. International studies, although scarce, such as those by Amaro (2016), Baldridge and Erk (2008), and Fosler-Lussier (2008), reveal challenges related to the aversion to formal representations and the heterogeneity of students' knowledge. The case of the University of Tartu (Koit et al., 2002) demonstrates a successful interdisciplinary model, which reinforces the importance of investigating the Brazilian context to improve the education of Letters and Linguistics students. The methodology adopted is a mixed method, based on content and document analysis (Moraes, 1999; Libâneo, 1994). Publicly available computational linguistics syllabi from Language and Literature courses at Brazilian universities were collected and systematized. The analysis focused on syllabi, objectives, methodologies, and bibliographical references, seeking pedagogical patterns and trends. The results indicate the presence of computational linguistics in 3 undergraduate and 5 graduate programs. Systematizing the data in a spreadsheet allowed us to identify the most frequent references and analyze the didactic and pedagogical nature of the programs. This initial survey suggests recognition of the field but highlights the need for a deeper understanding of teaching practices. The presence of the discipline in several institutions is a positive indicator, but the scarcity of specific literature on the subject and the complexity of the pedagogical challenges demand future research. Thus, this work serves as a basis for research aimed at improving curricula and education of linguists and other language professionals.

KEYWORDS: Computational Linguistics. Higher Education. Letters. Linguistics. Natural Language Processing.

Artigo recebido em: 04.11.2025
Artigo aprovado em: 31.01.2026

1 Introdução

De acordo com a página inicial do site da *Association for Computational Linguistics* (ACL), a “linguística computacional é o estudo científico da língua a partir de uma

perspectiva computacional”¹ (tradução nossa). Em geral, ela pretende gerar modelos computacionais² de vários tipos para os fenômenos linguísticos. Além disso, de acordo com o *site* da associação, o trabalho de linguistas computacionais pode envolver uma motivação científica (quando tais profissionais buscam dar uma explicação computacional para um fenômeno linguístico ou psicolinguístico) ou uma motivação puramente tecnológica (quando buscam desenvolver um componente que funcione dentro de um sistema de processamento de linguagem natural).

Dentro dessa perspectiva, vale ressaltar que as relações entre os elementos “computacionais” e “linguísticos” são variadas e apresentam a possibilidade de algumas interpretações, a começar pela própria nomenclatura da área.

Há autores que tratam os termos “linguística computacional” e “processamento de linguagem natural” como sinônimos; de outro lado, autores que estabelecem uma diferença, vinculando o processamento de linguagem natural ao campo da ciência da computação e a linguística computacional à área da linguística.

Apenas pela variação na nomenclatura da área pode-se notar que se trata de uma área interdisciplinar (Church; Liberman, 2021; Alencar, 2023; Freitas, 2022). Assim, é notável que um grande número de relações entre língua e computação venham à tona e apresentem características variadas e similares, formando certos padrões de linhas investigativas.

Em geral, na linguística computacional, seja teórica ou aplicada, há desenvolvimento em quatro linhas na interface entre a linguística e a ciência da computação (Alencar, 2023; Guinovart, 2000):

- 1) informática aplicada à pesquisa linguística;
- 2) implementação de teorias linguísticas;
- 3) aplicações linguísticas da informática;
- 4) simulação de processos cognitivos.

¹ Disponível em: <https://www.aclweb.org/portal/> . Acesso em: 03 nov. 2025.

² Tais modelos, de acordo, com o site da ACL (<https://www.aclweb.org/portal/>), podem ser do tipo “baseados em conhecimento” (“realizados manualmente”) ou “direcionados por dados” (“estatísticos” ou “empíricos”; tradução nossa).

É em uma ou mais de uma dessas quatro áreas que os linguistas computacionais situam suas pesquisas. Algumas universidades possuem graduação específica para tal formação (veja exemplos na próxima seção). Mas em geral, no contexto brasileiro, tal papel se dá fora (ou para além) de uma formação superior.

No Brasil, a linguística computacional, praticada, nos seus primórdios no país, quase exclusivamente por cientistas da computação, físicos e engenheiros, foi encampada institucionalmente pela Sociedade Brasileira de Computação (SBC), por meio da Comissão Especial de Processamento de Linguagem Natural (CE-PLN), aprovada no XXVII encontro daquela associação no Rio de Janeiro em 2007, graças à iniciativa das cientistas da computação Maria das Graças Volpe Nunes, da Universidade de São Paulo em São Carlos, e Renata Vieira e Vera Lúcia Strube de Lima, ambas da Pontifícia Universidade Católica do Rio Grande do Sul³. Consequentemente, muito do trabalho de formação (ainda que fragmentária) em tal área interdisciplinar acabou adotando o paradigma do processamento de linguagem natural e da ciência da computação. Ao longo dos anos, porém, essa comissão tem agregado cada vez mais linguistas, cuja participação tem aumentado no Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL), o principal evento que promove a área.

Do lado da área de Letras e Linguística, a linguística computacional ainda hoje não tem igual respaldo das agremiações representativas da área, como a Associação Nacional de Pós-graduação em Letras e Linguística (ANPOLL) e a Associação Brasileira de Linguística (ABRALIN). Mesmo assim, há que se destacar a divulgação da área, inclusive entre alunos de graduação, por meio de minicursos, mesas redondas e simpósios esporádicos, dos quais não podemos elencar senão alguns exemplos, a saber: (i) a Escola Brasileira de Linguística Computacional (EBRALC), evento satélite do tradicional Encontro de Linguística de Corpus (ELC), os quais, em 2024, realizaram a sua XIII e XVI edições, respectivamente; (ii) o Simpósio de Linguística

³ Disponível em: <https://sites.google.com/view/ce-pln/> . Acesso em: 03 nov. 2025.

Computacional do 68º Seminário do Grupo de Estudos Linguísticos do Estado de São Paulo (GEL), ocorrido em 2021 (Alencar; Lopes, 2021); (iii) minicurso de programação para linguistas e Simpósio Temático “Avanços da Linguística de Corpus e Computacional” por ocasião do XXIV Instituto de Linguística e XI Congresso Internacional da ABRALIN em Maceió em maio de 2019, que contou com a composição de uma Comissão Científica da área de Linguística de Corpus e Computacional, reunindo linguistas de diferentes regiões do Brasil; (iv) 1º Colóquio de Linguística para o Processamento Automático de Linguagem Natural (LiPrAL), realizado na UFES em 2012 com a participação de pesquisadores tanto da linguística quanto da ciência da computação (Laporte; Smarsaro; Vale, 2013).

A recente proliferação de tecnologias de inteligência artificial generativa destaca a necessidade cada vez mais urgente de integrar a linguística computacional na formação de profissionais de Letras. Uma das referências recentes que atestam tal necessidade é o recente artigo do linguista Tiago Torrent para o portal de notícias UOL (Torrent, 2025). O texto apresenta diversas críticas ao desenvolvimento indiscriminado de *Large Language Models* (LLMs) como parte essencial da elaboração do Plano Brasileiro de Inteligência Artificial (PBIA), defendendo a necessidade de inclusão de estudos linguísticos na discussão e elaboração de tal documento.

Alguns cursos de Letras oferecem disciplinas relacionadas à linguística computacional, contudo não se sabe o que é feito em sala de aula. Não se sabe como se dá o processo de ensino-aprendizado, ou quais perspectivas são adotadas. Quais materiais didáticos são utilizados em sala? Qual é o perfil do aluno de Letras ao realizar tal disciplina? Quais suas facilidades e desafios com base em sua realidade acadêmica?

Tais perguntas motivam a investigação sobre a questão didática da Linguística Computacional vista na literatura da área (veja seção subsequente), na demanda dos alunos de Letras em sala de aula, e nos desafios (tais como conhecimentos de elementos de lógica e matemática) na prática do ensino de linguística.

Nesse contexto, há uma escassez de investigação sobre o processo didático-pedagógico da linguística computacional, mais ainda, sobre o contexto específico da realidade do ensino superior brasileiro e dentro dos cursos de Letras.

Assim, busca-se, neste trabalho, analisar o perfil dos programas de disciplina de linguística computacional em cursos de graduação e pós-graduação em Letras de universidades brasileiras.

Este trabalho justifica-se especialmente pela necessidade de maior investigação com viés da educação na área de linguística computacional como um todo. Em específico, busca contribuir para a formação de estudantes da grande área de Letras e Linguística no contexto universitário brasileiro, ressaltando a importância de tal perfil em uma área que, embora multidisciplinar, demanda maior atenção e capacitação de docentes e discentes de Letras.

2 Pressupostos teóricos

Podemos constatar que não há, na literatura em contexto nacional, trabalhos que relatem o processo de ensino-aprendizagem de linguística computacional. O cenário internacional não é muito diferente, contando com poucas amostras, ainda que valiosas, como a de Amaro (2016, p. 40). De acordo com o seu trabalho, em cursos de bacharelado e mestrado de linguística computacional na área de linguística ou humanidades, há uma tendência em se utilizar abordagens sintáticas fortemente baseadas em formalizações (por exemplo, *Head-Driven Phrase Structure Grammar* - HPSG, *Lexical Functional Grammar* – LFG e gramáticas de dependências) em uma única disciplina, o que seria insuficiente para cobrir todas as diferentes subáreas, objetivos e tarefas da linguística computacional, além de uma noção básica de lógica e programação, o que permitiria uma experiência de aprendizado mais facilitada e sólida.

O aprendizado de formalismos gramaticais integra tradicionalmente a formação em linguística, por exemplo, sob a perspectiva de paradigmas como o

estruturalista e o gerativista. A formalização é praticada também no âmbito de algumas abordagens funcionalistas, como, por exemplo, Dik (1992). Paradoxalmente, contudo, Amaro (2016) afirma que o principal desafio para discentes de linguística e humanidades no aprendizado de linguística computacional parece ser [justamente] a “aversão” a representações formais. Tais modelos e representações utilizam símbolos e noções de lógica, de matemática e de ciência da computação, tal como predicados, quantificadores, operadores booleanos, teoria dos conjuntos, listas e outros elementos que tais estudantes não veem desde a (ou jamais vistos na) educação básica.

Todavia, Alencar (2009, p. 135) alerta para a necessidade de inclusão de elementos de diversas áreas, sobretudo da Ciência da Computação, na formação de estudantes e pesquisadores das áreas de Letras e Linguística:

O envolvimento com a linguística de corpus, por parte de estudantes e pesquisadores nas áreas de Letras e Linguística, será tão mais proveitoso quanto mais levar em conta o quadro interdisciplinar em que a disciplina está inserida, do qual faz parte também a linguística computacional.

Complementando Alencar (2009), Alencar e Rademaker (2022) adere à ideia da linguista computacional Emily Bender da implementação computacional como método de testagem de teorias e análises gramaticais (Bender 2002, 2003, 2010a, 2010b, 2011). Alencar e Rademaker (2022) se insere nessa linha, relacionando-se com a tendência apontada por Amaro (2016) de se utilizar abordagens gramaticais baseadas em formalização dentro do âmbito da linguística computacional no contexto da área de Letras e Linguística.

Já os trabalhos de Baldridge e Erk (2008) e Fosler-Lussier (Fosler-Lussier, 2008) tratam dos desafios de ensino de linguística computacional para um perfil de aula e de discentes heterogêneos, em especial do ponto de vista de conhecimentos dos fundamentos e da diversidade de formação dos alunos que participam dessa disciplina dentro do contexto acadêmico estadunidense. O primeiro trabalho se atém

especificamente à formação de uma disciplina interdepartamental, dada pela universidade, e o segundo enfoca o relato dos desafios e êxitos no planejamento e desenvolvimento das aulas para um público também heterogêneo de discentes, matriculados na disciplina.

Anterior a esses trabalhos está o caso excepcional do trabalho de Koit et al. (2002), que apresenta os resultados da implementação de um curso de bacharelado em linguística computacional na Universidade de Tartu, Estônia. Parte da formação é realizada no departamento de Filosofia e outra parte no departamento de Matemática da universidade, com possibilidade de extensão para mestrado na área de linguística computacional. As disciplinas contam com a presença de estudantes de outras áreas, além do alunado próprio do bacharelado em linguística computacional.

No contexto nacional, observemos parcerias de pesquisadores e seus orientandos de graduação e pós-graduação em Letras e Linguística na publicação de seus trabalhos atuando no desenvolvimento de recursos e ferramentas para as línguas indígenas brasileiras. Esse é o caso de Alexandre et al. (2021a; 2021b), Santos (2024), e Dias e Pires (2025), por exemplo.

Outro trabalho que associa pesquisa linguística com desenvolvimento de recursos e ferramentas computacionais, mas no contexto internacional, é o de Zariquiey et al. (2022), que descreve a construção de um *treebank* (*corpus* anotado sintaticamente) para a língua indígena Kakataibo, anotado por meio da abordagem *Universal Dependencies* (UD), dentro de um contexto colaborativo em uma disciplina de Linguística Computacional para graduandos. O trabalho se mostrou efetivo e alinhado com a abordagem (re)vitalizadora do repositório de recursos e ferramentas para línguas indígenas peruanas, o Churana.

A linguística computacional deve ser destacada como ferramenta potencial para a revitalização das línguas nacionais, pois a falta desse

suporte impede o crescimento dessas línguas e seu uso produtivo na Internet (e em qualquer sistema eletrônico) (tradução nossa)⁴.

O desenvolvimento de recursos, como *corpora* digitais e ferramentas computacionais para documentação e preservação das línguas indígenas, promove a valorização da diversidade linguística, o engajamento ético com as tecnologias e a articulação interdisciplinar entre linguística, ciência da computação e políticas linguísticas (Alencar, 2023, p. 78-79).

Assim, como relatado em Zariquiey *et al.* (2022), estudantes do curso bacharelado em Linguística Computacional da Universidade de Tübingen participam da elaboração de *treebanks* para diversas línguas, incluindo línguas originárias das Américas, por meio da abordagem UD. Esses estudantes desenvolvem habilidades linguísticas (por exemplo, formalismos gramaticais e linguística geral) e linguístico-computacionais (por exemplo, *parsing* e tecnologia do texto). Habilidades em programação não são um pré-requisito, pois são ensinadas ao longo do curso⁵.

Desde 2004, em 21 edições da disciplina "Ling 567" (*Knowledge Engineering for NLP*) da Universidade de Washington, Estados Unidos, ministrada por Emily Bender e associados, foram implementadas gramáticas de 165 línguas de 58 famílias linguísticas, incluindo algumas línguas indígenas brasileiras⁶. Esse curso une tipologia linguística, linguística computacional e engenharia do conhecimento. A implementação computacional da descrição gramatical de uma língua no sistema *Grammar Matrix* permite automaticamente gerar análises semânticas para as sentenças abarcadas por essa descrição. Permite, igualmente, testar a compatibilidade entre diferentes componentes de uma gramática, assegurando a sua coerência interna,

⁴ No original: *computational linguistics should be highlighted as the tool potential for the revitalization of national languages, as the lack of this support prevents the growth of these languages and their productive use on the Internet (and in any electronic system).* (Disponível em: <https://github.com/Llamacha/Churana> . Acesso em: 23 out. 2025)

⁵ Disponível em: <https://uni-tuebingen.de> . Acesso em: 24 out. 2025.

⁶ Disponível em: <https://courses.washington.edu/ling567/> e <https://wiki.ling.washington.edu>. Acesso em: 23 out. 2025.

mesmo no caso de gramáticas bastante extensas e complexas, algo difícil de verificar manualmente. Por outro lado, essa abordagem permite verificar automaticamente que sentenças a gramática abrange, determinando, ao mesmo tempo, que construções violam os princípios implementados na gramática⁷.

É com base nesse contexto e a partir dessas limitações que se faz mister levantar informações sobre a situação do processo de ensino-aprendizagem de linguística computacional no contexto brasileiro. Para que os achados possam, de certa forma, incluir avanços na formação de estudantes de Letras como futuros profissionais de linguística computacional.

3 Metodologia

O presente trabalho realiza um levantamento dos programas/planos de ensino de disciplinas de linguística computacional em cursos de graduação e pós-graduação de Letras e/ou Linguística no Brasil. Tal levantamento tem como objetivo apontar padrões no uso das referências bibliográficas, ementas e demais itens informativos sobre o caráter da disciplina. Em seguida, realiza uma breve revisão das principais referências identificadas no levantamento. Por fim, tece uma análise do perfil didático-pedagógico por meio do cruzamento das etapas anteriores.

Para tanto, o presente estudo adota uma abordagem qualitativa fundamentada na análise de conteúdo e na análise de documentos, conforme delineado por Moraes (1999). Ambas as metodologias são apropriadas para estudos voltados à compreensão de significados e estruturas latentes em materiais textuais e documentais, permitindo interpretar os dados à luz de contextos específicos, neste caso, do ensino da linguística computacional em cursos de Letras no Brasil.

A análise de conteúdo, segundo Moraes (1999), é um conjunto de técnicas metodológicas que visa à interpretação de textos a partir de procedimentos

⁷ Para mais detalhes, veja Alencar e Rademaker (2022).

sistemáticos e objetivos de descrição e inferência. Essa metodologia possibilita a categorização dos dados em unidades temáticas e a identificação de padrões recorrentes, permitindo compreender tanto os conteúdos explícitos quanto os implícitos nas fontes analisadas. No contexto deste estudo, a análise de conteúdo foi aplicada na interpretação das ementas, objetivos, metodologias e referências bibliográficas dos programas de ensino das disciplinas de linguística computacional.

Complementarmente, utilizou-se a análise de documentos, também conforme Moraes (1999), entendida como o exame sistemático de materiais que não foram produzidos para fins de pesquisa, mas que são ricos em informações relevantes sobre o objeto de estudo. Tais documentos são compreendidos como testemunhos indiretos de práticas institucionais, revelando aspectos da cultura, da organização e das intencionalidades educativas presentes nos cursos de Letras. No caso deste trabalho, os documentos analisados foram os programas de ensino (ou planos de ensino) disponibilizados publicamente por universidades brasileiras em seus *sites* institucionais.

Para os fins desta investigação, define-se programa de ensino como o documento que organiza e apresenta formalmente os elementos estruturantes de uma disciplina acadêmica. De acordo com Libâneo (1994), o programa de ensino comprehende componentes como: título da disciplina, ementa, objetivos gerais e específicos, conteúdos programáticos, metodologias de ensino, formas de avaliação e bibliografia básica e complementar. Esses elementos articulam-se para orientar tanto o docente na condução do processo educativo quanto o discente em sua trajetória formativa. Assim, o programa de ensino é também uma expressão da intencionalidade pedagógica da instituição e um reflexo das concepções de ensino e aprendizagem que nela circulam.

Com base nesse referencial, o presente trabalho operacionalizou o levantamento em três etapas principais:

- 1) Coleta de dados documentais: realização de buscas no motor de pesquisa

Google com os termos “linguística computacional”, “processamento de língua natural”, “disciplina”, “ementa” e “programa”, de forma isolada ou combinada, com o intuito de localizar programas de ensino disponíveis publicamente em cursos de graduação e pós-graduação em Letras ou áreas afins em universidades brasileiras.

- 2) Sistematização e organização dos dados: preenchimento de uma planilha⁸ com os dados extraídos de cada programa identificado, incluindo instituição, nível do curso (graduação ou pós), elementos presentes (título, ementa, objetivos, metodologia, bibliografia) e, especificamente, as obras citadas nas referências.
- 3) Análise de conteúdo dos documentos: aplicação dos princípios da análise de conteúdo (Moraes, 1999), buscando compreender como os elementos pedagógicos estão estruturados nos programas, qual o perfil das bibliografias adotadas e quais tendências metodológicas e epistemológicas emergem dessas escolhas.

A escolha dessas metodologias justifica-se pela natureza documental da fonte primária deste estudo e pela necessidade de interpretar os significados presentes nas práticas de ensino da linguística computacional, prática ainda pouco explorada no contexto das Letras. Ao se basear em Moraes (1999), Libâneo (1994) e nos próprios documentos institucionais, este trabalho busca oferecer um retrato interpretativo do estado atual dessa disciplina na formação de estudantes de Letras, contribuindo para o avanço das discussões sobre a intersecção entre linguística computacional e ensino.

4 Resultados

Resultados do levantamento

O levantamento apresenta um total de 8 cursos entre graduação e pós-graduação em Letras/Linguística de 7 universidades no Brasil. Dos cursos de graduação estão a USP, UFSC e UFPB; e de pós-graduação, UFC, PUC-Rio, USP,

⁸ Ambas as planilhas se encontram disponíveis no repositório: <https://github.com/blanchunb>. Acesso em: 22. jan. 2026.

UFSCar e UFMG. O levantamento teve como base a disponibilização na internet de programas de ensino de disciplina de linguística computacional ou área afim que demonstrassem em suas referências ao menos uma obra relacionada à linguística computacional.

O número aparentemente pequeno reflete uma significativa porção dos cursos de Letras/Linguística que, embora disponibilizem informação de que há disciplina de linguística computacional ou com título relacionado a ela, não disponibilizam publicamente programa de disciplina para verificação das referências, título, ementa e demais informações necessárias para o levantamento dos dados deste trabalho. Esses são o caso, por exemplo, da disciplina “Língua e Programação” (Graduação – Línguas Estrangeiras Aplicadas, UnB) e “Linguística Computacional” (Graduação e Pós-Graduação em Linguística, Unicamp), além de diversas outras que podem constituir, eventualmente, disciplinas de “Tópicos Especiais” e nomes similares. Tais casos portanto, não foram levados em consideração para o cômputo dos resultados.

Feita essa ressalva, pelo levantamento das oito ocorrências de programas, foi possível observar que a maioria dos cursos se concentra na região sudeste (5), especialmente no estado de São Paulo (3). Há ocorrências, contudo, na região nordeste (2) e sul do país (1).

Uma relativa maioria desses cursos na pós-graduação (5) do que de graduação (3) demonstra uma aparente complexidade da linguística computacional para a graduação (onde figura majoritariamente em disciplinas optativas).

Há uma significativa heterogeneidade de referências bibliográficas entre os planos de ensino. A recorrência de referências bibliográficas entre os documentos encontrados segue a tendência internacional e da área da ciência da computação, repercutindo as obras *Speech and Language Processing* (Jurafsky; Martin, 2008), *Foundations of Statistical Language Processing* (Manning; Schütze, 1999) e *Natural Language Processing with Python* (Bird; Klein; Loper, 2008), com o cômputo do primeiro, de quatro vezes e dos demais, três vezes respectivamente. Além de tais obras, houve

ainda recorrência de *The Oxford Handbook of Computational Linguistics* (Mitkov, 2005) e da brasileira *Linguística Computacional: teoria & prática* de Othero e Menuzzi (2005).

A Tabela 1 a seguir apresenta a recorrência dessas obras, o que delimita ainda mais o escopo dos cursos para seis (6), e apenas cinco (5) instituições, onde o curso de graduação da UFSC e de pós-graduação da UFMG não apresentam em seus planos de ensino referência bibliográfica comum ao padrão de outras instituições. Isso acontece na medida em que se direciona a linguística computacional enquanto complemento, método, ou ferramenta, ou então, o plano de ensino apresenta bibliografia significativamente datada, como é o caso da UFSC (referências do início dos anos 90 e final dos 80, sendo o plano de ensino, contudo, apresentado para o semestre de 2021.1).

Tabela 1 – Recorrência de obras por programa de ensino.

Referência bibliográfica	Ocorrência em programas	Cursos/Universidades
JURAFSKY, Daniel; MARTIN, James. H. Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. Upper Saddle River, NJ: Prentice Hall, 2008.	4	<ul style="list-style-type: none"> • Pós-Graduação em Linguística da UFC, • Pós-Graduação em Linguística da USP, • Graduação em Linguística da USP, • Pós-Graduação em Linguística da UFSCAR.
MANNING, C. e SCHÜTZE, H. Foundations of Statistical natural language processing. Cambridge, MA: The MIT Press, 1999.	3	<ul style="list-style-type: none"> • Pós-Graduação em Estudos da Linguagem da PUC-Rio, • Pós-Graduação em Linguística da USP, • Graduação em Linguística da USP,
BIRD, Steven; KLEIN, Ewan; LOPER, Edward. Natural Language Processing in Python. Sebastopol: O'Reilly Media, 2008.	3	<ul style="list-style-type: none"> • Pós-Graduação em Linguística da UFC, • Pós-Graduação em Linguística da USP, • Graduação em Linguística da USP,
MITKOV, Ruslan. The Oxford handbook of computational	3	<ul style="list-style-type: none"> • Pós-Graduação em Linguística da UFC; • Pós-Graduação em

linguistics. Oxford: Oxford University Press, 2005.

Estudos da Linguagem da PUC-Rio,

- Pós-Graduação em Linguística da UFSCAR,

OTHERO, Gabriel de Ávila; MENUZZI, Sérgio de Moura.

3

- Pós-Graduação em Linguística da UFC,
- Pós-Graduação em Linguística da UFSCAR,
- Graduação em Letras-Português da UFPB

Lingüística Computacional: teoria & prática. São Paulo: Parábola, 2005.

Fonte: elaborada pelos autores.

É importante salientar que, entre as documentações encontradas, a maioria daquelas que apresentam data são da década passada, o que reflete, de certa forma a baixa ocorrência de obras recentes e relevantes no cenário nacional (veja subseção "Revisão das principais referências"). Ainda assim, é notável haver uma referência em língua portuguesa dentro do programa de ensino nos cursos da UFPB, UFSCAR e UFC.

Embora seja verdade que uma grande parte dos planos de ensino/ programas não foram levados em consideração⁹, ainda são documentações representativas de cada instituição, uma vez que devem disponibilizar o material atualizado para avaliações institucionais periódicas e facilitar o acesso a informações didático-pedagógicas.

A partir dos seis programas de disciplinas levantados e apresentados na Tabela 1, foi possível identificar a presença ou ausência de certos componentes estruturais nos documentos analisados. Os principais elementos observados foram: título da disciplina, ementa, objetivos, conteúdo programático, metodologia de ensino,

⁹ Muitos desses documentos não são disponibilizados de forma online e outros tantos têm seu conteúdo atualizado informalmente a cada semestre.

avaliação e referências bibliográficas. O Quadro 1 a seguir resume a presença desses elementos por instituição.

Quadro 1 – Presença de componentes pedagógicos nos planos de ensino analisados.

Universidade	Nível	Ementa	Objetivos	Metodologia	Avaliação	Conteúdo Programático	Referências
UFC – Pós-graduação	Pós	Sim	Sim	Parcial	Não	Sim	Sim
USP – Pós-graduação	Pós	Sim	Sim	Sim	Não	Sim	Sim
USP – Graduação	Graduação	Sim	Parcial	Sim	Sim	Sim	Sim
UFSCAR – Pós-graduação	Pós	Sim	Sim	Sim	Não	Parcial	Sim
PUC-Rio – Pós-graduação	Pós	Sim	Parcial	Não	Não	Sim	Sim
UFPB – Graduação	Graduação	Sim	Sim	Parcial	Parcial	Sim	Sim

Fonte: elaborado pelos autores.

Ao observar os dados do Quadro 1, emergem alguns padrões e inferências importantes. Ementa e referências são os únicos elementos universalmente presentes, sugerindo uma preocupação comum com o enquadramento temático das disciplinas e com o suporte bibliográfico. No entanto, as referências bibliográficas são predominantemente estrangeiras, com poucas obras em português.

Objetivos aparecem em todos os planos, mas com níveis variados de formulação. Aqueles marcados como “parcial”, por exemplo, aparecem no documento, mas apenas de modo reduzido ou genérico, sem detalhamento explícito.

A metodologia de ensino é o elemento mais frágil entre os planos analisados. Em apenas um caso (USP – pós), ela aparece detalhadamente. Nos demais, ou está ausente ou descrita de maneira vaga, com expressões como “aulas expositivas” ou “discussão de textos”, sem indicação clara de estratégias específicas, recursos tecnológicos, ou formas de mediação ativa dos conteúdos.

A avaliação é o componente mais negligenciado. Nenhum plano de pós-graduação explicita critérios avaliativos e apenas um de graduação (USP) descreve

claramente os métodos de avaliação adotados. Isso indica uma lacuna significativa em termos de transparência e acompanhamento do processo de aprendizagem.

O conteúdo programático está presente em todos os planos, e, embora com variações, abrange geralmente temas como linguagens formais, processamento de linguagem natural, ferramentas computacionais e representação linguística. Em alguns casos, percebe-se uma tentativa de equilibrar teoria e prática; em outros, a abordagem parece mais técnica ou voltada a tópicos de pesquisa específicos.

Predomina a ausência de articulação didático-pedagógica explícita. Os elementos dos planos não são apresentados de forma integrada: raramente há conexão entre objetivos, metodologia e avaliação, o que dificulta a leitura de um projeto pedagógico coeso.

Ainda que haja alguma convergência nas obras de referência como Jurafsky e Martin (2008) e Bird *et al.* (2008), além de certa corroboração de bibliografias de cursos de processamento de linguagem natural (na ciência da computação) em contexto nacional e internacional, o formato e a organização dos planos são bastante distintos, o que evidencia a ausência de diretrizes curriculares comuns para a área nos cursos de Letras.

A título de curiosidade e de exceção à regra, em um curso de pós-graduação em matemática aplicada e ciência de dados da FGV-EMAp, há uma disciplina eletiva de “Introdução à Linguística” e outra de “Processamento de Linguagem Natural e Mineração de textos”¹⁰. A primeira corrobora, curiosamente, em sua bibliografia exatamente a referência mais citada aqui, a de Jurafsky e Martin (2008), além de outras referências da linguística de corpus. Já a segunda não inclui essa referência, mas a de Manning e Schütze (1999), entre outras relacionadas à métodos simbólicos e estatísticos aplicados aos estudos da linguagem. Sobretudo, é inusitado observar a

¹⁰ Disponível em: <https://emap.fgv.br/curso/mestrado-em-matematica-aplicada-e-ciencia-de-dados#disciplinas>. Acesso em: 4 nov. 2025.

aplicação de abordagens linguísticas em curso de exatas, constatando sua importância para a ciência de dados e matemática aplicada.

Por fim, embora a maioria das disciplinas reconheça a natureza interdisciplinar da linguística computacional, poucas detalham como essa interdisciplinaridade se materializa na prática pedagógica — seja por meio de co-docência, atividades integradas ou parcerias com outras áreas.

A próxima etapa faz uma revisão das principais referências encontradas nesse levantamento, levando em consideração os métodos de análise de conteúdo e de documentos.

Revisão das principais referências

A obra *Speech and Language Processing*, de Jurafsky e Martin (2008), apresenta um marco fundamental na área de linguística computacional, especialmente no contexto do ensino da disciplina em ambientes interdisciplinares. Estruturada em diversas partes, a obra cobre desde fundamentos linguísticos (como fonética e morfologia computacional), até aspectos estatísticos, como modelos de linguagem e aprendizado de máquina. Sua abordagem equilibra teoria linguística e prática computacional, sendo indicada tanto para estudantes de ciência da computação quanto de linguística. No entanto, exige certo domínio prévio de matemática, estatística e lógica, o que pode representar um desafio para discentes de Letras que não possuam conhecimento nessas áreas.

Em *Foundations of Statistical Natural Language Processing*, de Manning e Schütze (1999), observa-se uma ênfase ainda mais pronunciada na estatística aplicada ao processamento de linguagem. A obra se destaca por apresentar formalizações matemáticas dos modelos probabilísticos empregados em tarefas como análise sintática (*parsing*), etiquetamento morfossintático (*PoS Tagging*) e modelagem de tópicos. Assim como o livro de Jurafsky e Martin (2008), seu uso em cursos de Letras pode ser dificultado pela necessidade de conhecimentos prévios em cálculo e

estatística, embora ofereça explicações detalhadas que podem ser adaptadas para públicos diversos mediante mediação didática apropriada.

Já *Natural Language Processing with Python*, de Bird, Klein e Loper (2008), se diferencia por adotar uma abordagem prática e introdutória voltada ao uso da biblioteca NLTK (*Natural Language Toolkit*). A obra permite que os estudantes interajam com os conceitos por meio da programação, em especial com a linguagem *Python*, acessível e amplamente utilizada na área. Sua orientação aplicada, aliada a uma curva de aprendizagem relativamente suave, faz dessa obra uma excelente porta de entrada para estudantes de Letras interessados em desenvolver e aplicar na prática algumas tarefas de processamento de linguagem natural, embora ainda demande um esforço de ambientação com a linguagem de programação.

The Oxford Handbook of Computational Linguistics, organizado por Mitkov (2005), reúne textos de diferentes autores abordando múltiplas dimensões da área, incluindo aspectos linguísticos, tecnológicos e históricos. Seu caráter abrangente e crítico faz da obra uma referência importante para estudantes de pós-graduação, principalmente aqueles que desejam obter uma visão panorâmica das subáreas da linguística computacional. Por sua densidade e escopo, tende a ser mais apropriada para cursos em nível de pós-graduação do que para disciplinas introdutórias.

Por fim, *Linguística Computacional: teoria & prática*, de Othero e Menuzzi (2005), representa uma das poucas referências nacionais recorrentes, escrita em língua portuguesa e pensada para o contexto brasileiro. A proposta dos autores é mostrar como o computador pode ser usado como instrumento para a descrição, modelagem e simulação da linguagem natural, abordando tanto aspectos teóricos quanto aplicações práticas. Emprega abordagens simbólicas e lógicas, sem discutir métodos estatísticos ou de aprendizado de máquina que predominam na área atualmente. Além disso, utiliza a linguagem de programação *Prolog*, mas não *Python* como tem sido o costume na última década.

É importante enfatizar que essa última obra foi a única referência brasileira citada em mais de um plano de ensino. Isso é devido, entre outros motivos, à criação já datada dos planos, o que tornaria a obra didática de Othero e Menuzzi (2005) a principal em língua portuguesa disponível no mercado editorial brasileiro durante quase quinze anos, até a publicação de Ferreira e Lopes (2019).

Freitas (2022) apresenta uma breve introdução teórica e prática à linguística computacional, utilizando-se de uma linguagem mais acessível ao público de Letras/Linguística, abrangendo diversas áreas da linguística computacional, e alocada em uma coleção tipicamente sobre áreas da linguística. Essa é talvez a obra que mais busca delimitar o papel e nicho de linguistas dentro da área da linguística computacional. A autora destaca que a linguística computacional deve ser encarada, sobretudo, sob a ótica linguística e não meramente como computação, ou seja, devemos questionar as suas implicações para a linguagem, para o uso social e para a educação. A obra dedica o capítulo mais extenso à tarefa que considera, de fato, linguística, dentro da linguística computacional: a anotação de *corpus*. O livro disponibiliza alguns exercícios práticos e referências para aprofundamento em cada tema.

Caseli e Nunes (2022) apresentam um livro digital sobre o processamento de língua natural, com diversos exemplos, explicações e conceitos de muitas subáreas, aplicações e tarefas de linguística computacional. Contudo, oferecem um viés a partir do Processamento de Linguagem Natural (PLN), com explicações técnicas e teóricas acessíveis para o público geral, mas sem exercícios de fixação.

A obra que talvez proponha uma estrutura mais próxima a de um curso de linguística computacional, funcionando como base para ensino em sala de aula, é a de Ferreira e Lopes (2019). A obra, no entanto, é altamente seletiva e deixa fora de seus capítulos uma série de temas relevantes e reflexivos para a área em função da proposta traçada. Entende também, assim como Caseli e Nunes (2022), a linguística computacional pelo viés do processamento de língua natural.

Apesar de essas últimas referências não figurarem nos programas de ensino levantados e apresentarem propostas com limitações para a formação do público aqui caracterizado, tais referências são, ainda assim, de grande relevância para a área no contexto de nacional.

Análise do perfil didático-pedagógico

A partir do levantamento de planos de ensino e revisão das referências bibliográficas, é possível delinejar um panorama crítico do ensino de linguística computacional no Brasil, com base nas Diretrizes Curriculares Nacionais para os cursos de Letras (Resolução CNE/CES nº 1/2002). Tais diretrizes apontam para uma formação sólida em estudos linguísticos, literários e culturais, com ênfase na leitura, interpretação, produção textual e reflexão crítica. No entanto, os achados do presente estudo indicam que a estrutura pedagógica dos cursos analisados encontra dificuldades em integrar essa formação humanística aos saberes técnicos exigidos pela linguística computacional.

O cruzamento entre os dados dos planos de ensino e o perfil formativo dos estudantes revela três frentes de desalinhamento. Primeiramente, observa-se uma disparidade formativa entre áreas: estudantes de Letras, por sua trajetória centrada em abordagens teórico-analíticas da linguagem, costumam demonstrar resistência ou dificuldade com representações formais e abstrações matemáticas, que são elementos basilares da linguística computacional. Como apontam Amaro (2016) e Baldridge e Erk (2008), esse descompasso se acentua quando não há estratégias de transposição didática adequadas.

Em segundo lugar, constata-se ausência de mediação pedagógica planejada. A maioria dos planos de ensino analisados não apresenta metodologias claras, e raramente há conexão explícita entre objetivos, conteúdos e formas de avaliação. A disciplina, portanto, do ponto de vista da análise documental, tende a ser conduzida

de modo técnico-expositivo, sem articulação com estratégias formativas ou recursos que considerem o perfil específico do público-alvo.

Por fim, destaca-se o desalinhamento entre o material didático adotado e o perfil discente. As principais obras utilizadas nos programas analisados (como Jurafsky; Martin, 2008; Manning; Schütze, 1999; Bird *et al.*, 2008) são redigidas em inglês técnico e exigem conhecimentos prévios em lógica, estatística e ciência da computação. Uma única obra nacional foi recorrente, qual seja, a obra de Othero e Menuzzi (2005), deixando obras mais recentes como Ferreira e Lopes (2019) e Freitas (2022) de fora, o que reforça a barreira linguística e cognitiva para estudantes da graduação em Letras.

Além desses desafios estruturais, destaca-se que a disciplina de linguística computacional ainda é frequentemente tratada como um “complemento” metodológico optativo, especialmente nos cursos de graduação. Essa posição periférica compromete sua eficácia como espaço de aprendizagem crítica e interdisciplinar, além de restringir a possibilidade de os estudantes refletirem, de forma efetiva, sobre o papel da linguagem nas tecnologias contemporâneas.

5 Considerações finais

Este estudo teve como objetivo principal identificar o perfil do processo de ensino-aprendizagem de linguística computacional no contexto de cursos de Letras e Linguística no Brasil, por meio da análise de programas de disciplinas, referências bibliográficas e metodologias de ensino empregadas. O levantamento documental e a análise de conteúdo permitiram compreender como essa área, de natureza profundamente interdisciplinar, tem sido inserida (ou negligenciada) nas estruturas curriculares dessas formações.

Entre os principais achados, destaca-se a escassez de programas de ensino disponíveis publicamente, o que já indica um primeiro obstáculo à transparência e à sistematização da disciplina. Os planos de ensino analisados apresentaram

significativa heterogeneidade em estrutura e profundidade, comementas e bibliografias amplamente baseadas em obras estrangeiras e tecnicamente exigentes. Foi observada ainda a ausência ou fragilidade de componentes essenciais como metodologias de ensino e critérios de avaliação, o que aponta para uma baixa formalização didático-pedagógica. Também se evidenciou o descompasso entre os materiais didáticos utilizados e o perfil dos estudantes de Letras, com raríssimos exemplos de materiais produzidos em português ou adaptados às especificidades formativas desse público.

Outro achado importante diz respeito à marginalização da linguística computacional nos cursos de Letras, que aparece majoritariamente como disciplina optativa, técnica ou acessória, e raramente como eixo estruturante da formação. Isso colabora para que sua apropriação pelos discentes se mantenha limitada e distante de perspectivas críticas, inclusivas e contextualizadas.

Nesse cenário, duas frentes de atuação emergem como especialmente promissoras para preencher as lacunas apontadas. A primeira é a elaboração de materiais didáticos em português, especificamente voltados para estudantes de Letras, que apresentem os conceitos fundamentais da linguística computacional de forma acessível, progressiva e contextualizada. Tais materiais devem incluir glossários, atividades baseadas em problemas reais de linguagem e exemplos conectados às áreas tradicionais da formação em Letras, como morfossintaxe, ensino de línguas, literatura e análise do discurso, por exemplo.

A segunda frente consiste no desenvolvimento de ferramentas e recursos computacionais voltados para línguas indígenas brasileiras. Essa iniciativa, além de responder a uma demanda científica e política de valorização da diversidade linguística nacional, representa uma oportunidade pedagógica singular. Ao aplicar conhecimentos da linguística computacional a línguas historicamente marginalizadas nos currículos universitários, o estudante desenvolve competências técnicas e, ao mesmo tempo, se engaja com uma perspectiva crítica, ética e socialmente situada.

Projetos de documentação, construção de corpora, modelagem gramatical e anotação morfossintática de línguas indígenas são exemplos de práticas que conectam teoria, tecnologia e política linguística e que contribuem para ampliar o escopo formativo da disciplina (Alencar, 2023, p. 78-79).

Dessa forma, a análise aqui realizada aponta não apenas para os desafios estruturais do ensino de linguística computacional em Letras, mas também para possibilidades concretas de reconfiguração pedagógica. A partir da produção de materiais adaptados, da valorização da diversidade linguística brasileira e do fortalecimento da interdisciplinaridade, é possível avançar em direção a uma formação linguística que dialogue com as demandas tecnológicas, sociais e políticas do presente.

Entre os principais limites da pesquisa, destaca-se a dependência de documentos publicamente acessíveis, o que restringe o escopo do levantamento. Muitos planos de ensino atualizados são disponibilizados apenas internamente pelas instituições ou variam a cada semestre sem registro público contínuo. Além disso, o estudo se concentrou na análise documental, sem coleta de dados em campo (como entrevistas com docentes ou análise de práticas de sala de aula), o que poderia oferecer uma compreensão mais aprofundada das dinâmicas pedagógicas reais.

Com base nesses limites, propõem-se os seguintes caminhos para trabalhos futuros:

- Levantamentos empíricos com estudantes e docentes da área para investigar práticas didáticas, percepções, dificuldades e estratégias adotadas;
- Desenvolvimento, aplicação e avaliação de materiais didáticos experimentais em cursos de Letras;
- Mapeamento e acompanhamento de projetos de linguística computacional voltados para línguas indígenas, com vistas à produção de modelos pedagógicos replicáveis;
- Construção colaborativa de diretrizes curriculares e referenciais pedagógicos nacionais para o ensino de linguística computacional em Letras, promovendo a consolidação da área como componente formativo crítico.

Avançar nessas direções permitirá não apenas fortalecer a presença da linguística computacional nos currículos de Letras, mas também democratizar o acesso ao letramento digital linguístico e aproximar os estudantes de Letras das discussões contemporâneas sobre linguagem, ciência e tecnologia.

Referências

ACL. **Association for Computational Linguistics**, 2025. What is computational linguistics? Disponível em: <https://www.aclweb.org/portal/>. Acesso em: 3 nov. 2025.

ALENCAR, L. F. de. Técnicas em software livre para exploração de corpora do português livremente disponíveis na WWW. **Veredas - Revista de Estudos Linguísticos**, v. 13, n. 2, 2009.

ALENCAR, L. F. de; LOPES, M. Simpósio de Convidados 11/13: Linguística Computacional. **68º Seminário do GEL**. Disponível em: https://www.gel.org.br/upload/arquivos/20210114_39_resumos-68-seminario/142-simposio-11.pdf. Acesso em: 22 jan. 2026.

ALEXANDRE, D. M.; GURGEL, J. L.; ARARIPE, L. F. de A. Nheentquetador: um etiquetador morfossintático para o sintagma nominal do nheengatu. **Revista Encontros Universitários da UFC**, v. 6, p. 1–13, 2021a.

ALENCAR, L. F. de; RADEMAKER, A. Modelação da valência verbal numa gramática computacional do português no formalismo HPSG. **Domínios de Lingu@gem**, v. 16, n. 4, p. 1339–1400, 12 set. 2022. DOI <https://doi.org/10.14393/DL52-v16n4a2022-6>

ALENCAR, L. F. de. Linguística Computacional. In: OTHERO, G. de Á.; FLORES, V. do N. (org.). **A linguística hoje: Múltiplos domínios**. São Paulo: Contexto, 2023.

ALEXANDRE, D. M.; GURGEL, J. L.; ARARIPE, L. F. de A. Compilação de um corpus etiquetado da Língua Geral Amazônica. In: **SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA (STIL)**, 13., 2021b, Evento online. **Anais** [...]. Porto Alegre: SBC, 2021b. p. 427–431. DOI <https://doi.org/10.5753/stil.2021.17823>

AMARO, R. Teaching computational linguistics: challenges and target audiences. In: **II WORLD CONGRESS ON COMPUTER SCIENCE, ENGINEERING AND TECHNOLOGY EDUCATION. Anais**[...]. Castelo Branco, Portugal: 2016.

BALDRIDGE, J.; ERK, K. Teaching computational linguistics to a heterogeneous student body: lessons from a graduate course. *In: PROCEEDINGS OF THE THIRD ACL WORKSHOP ON ISSUES IN TEACHING COMPUTATIONAL LINGUISTICS. Anais[...]* p. 1–8, 2008. DOI <https://doi.org/10.3115/1627306.1627308>

BENDER, E. M.; FLICKINGER, D.; OEPEN, S. The Grammar Matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. *In: COLING-GEE '02: Proceedings of the 2002 Workshop on Grammar Engineering and Evaluation.* [S.l.]: [s.n.], 2002. p. 8–14. DOI <https://doi.org/10.3115/1118783.1118785>

BENDER, E. M.; FLICKINGER, D.; OEPEN, S. **MRS in the LinGO Grammar Matrix: A practical user's guide.** [S.l.]: [s.n.], 2003. Disponível em: <http://faculty.washington.edu/ebender/papers/userguide.pdf>. Acesso em: 25 set. 2021.

BENDER, E. M. Reweaving a grammar for Wambaya: A case study in grammar engineering for linguistic hypothesis testing. *Linguistic Issues in Language Technology*, v. 3, p. 1–34, 2010a. DOI <https://doi.org/10.33011/lilt.v3i.1215>

BENDER, E. M. *et al.* Grammar customization. *Research on Language & Computation*, v. 8, n. 1, p. 23–72, 2010b. DOI <https://doi.org/10.1007/s11168-010-9070-1>

BENDER, E. M.; FLICKINGER, D.; OEPEN, S. Grammar engineering and linguistic hypothesis testing: Computational support for complexity in syntactic analysis. *In: BENDER, E. M.; ARNOLD, J. E. (ed.). Language from a cognitive perspective: Grammar, usage and processing.* Stanford: CSLI, 2011. p. 5–29.

BIRD, S.; KLEIN, E.; LOPER, E. **Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit.** Sebastopol: O'Reilly Media, 2009.

BRASIL. Ministério da Educação. Conselho Nacional de Educação. Câmara de Educação Superior. **Resolução CNE/CES nº 18, de 13 de março de 2002.** Brasília, DF, 2002.

CASELI, H. de M.; NUNES, M. das G. V. (org.). **Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português.** 1. ed. [s.l.] SBC, 2022. DOI <https://doi.org/10.5753/sbc.10309.7.1>

CE-PLN. **Comissão Especial de Processamento de Linguagem Natural**, 2025. Sobre a CE-PLN. Disponível em: <https://sites.google.com/view/ce-pln/>. Acesso em: 3 nov. 2025.

CHURANA. **Churana: Revitalizing Languages**, 2021. README. Disponível em: <https://github.com/Llamacha/Churana>. Acesso em: 4 nov. 2025.

CHURCH, K.; LIBERMAN, M. The Future of Computational Linguistics: On Beyond Alchemy. **Frontiers in Artificial Intelligence**, v. 4, p. 625341, 19 abr. 2021. DOI <https://doi.org/10.3389/frai.2021.625341>

DIAS, I. N.; PIRES, T. B. Aspectos do desenvolvimento de um etiquetador morfossintático da língua Asuriní do Trocará. *In: SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA (STIL)*, 16., 2025, Fortaleza/CE. **Anais** [...]. Porto Alegre: Sociedade Brasileira de Computação, 2025 . p. 618-626. DOI <https://doi.org/10.5753/stil.2025.37864>

DIK, S. C. **Functional Grammar in Prolog**: an integrated implementation for English, French, and Dutch. Berlin: Mouton de Gruyter, 1992. (Natural Language Processing, v. 2). 264 p. DOI <https://doi.org/10.1515/9783110850451>

FERREIRA, M.; LOPES, M. **Linguística Computacional**. São Paulo: Contexto, 2019.

FGV-EMAP. **Mestrado em Matemática Aplicada em Ciência de Dados**, 2025. Disciplinas Eletivas. Disponível em: <https://emap.fgv.br/curso/mestrado-em-matematica-aplicada-e-ciencia-de-dados#disciplinas> . Acesso em: 4 nov. 2025.

FOSLER-LUSSIER, E. Strategies for teaching “mixed” computational linguistics classes. Proceedings of the Third Workshop on Issues in Teaching Computational Linguistics - TeachCL '08. **Anais**[...]. *In: THE THIRD WORKSHOP*. Columbus, Ohio: Association for Computational Linguistics, 2008. DOI <https://doi.org/10.3115/1627306.1627313>

FREITAS, C. **Linguística Computacional**. São Paulo: Parábola, 2022.

GUINOVART, X. G.. Linguística Computacional. *In: RAMALLO, F.; REI-DOVAL, G.; RODRÍGUEZ, X. P. (org.). Manual de ciencias da linguaxe*. Vigo: Xerais, 2000. p. 221–268.

JURAFSKY, D.; MARTIN, J. H. **Speech and language processing**: An introduction to natural language processing, computational linguistics, and speech recognition. Upper Saddle River, N. J.: Prentice Hall., 2008.

KOIT, M. *et al.* Teaching Computational Linguistics at the University of Tartu. *In: PROCEEDINGS OF THE COLING WORKSHOP ON TEACHING*

COMPUTATIONAL LINGUISTICS. *Anais*[...], 2002. DOI <https://doi.org/10.3115/1118108.1118120>

LAPORTE, E.; SMARSARO, A.; VALE, O. A. (org.). **Dialogar é preciso: linguística para o processamento de línguas**. Vitória: PPGEL/UFES, 2013.

LIBÂNEO, J. C.. **Didática**. São Paulo: Cortez, 1994.

MANNING, C.; SCHÜTZE, H. **Foundations of Statistical Natural Language Processing**. Cambridge, MA: MIT Press, 1999.

MITKOV, R. (ed.). **The Oxford Handbook of Computational Linguistics**. Oxford: Oxford University Press, 2005.

MORAES, R. Análise de conteúdo. **Educação**, Porto Alegre, v. 22, n. 37, p. 7–32, jan./abr. 1999.

OTHERO, G. de Á.; MENUZZI, Sérgio de Moura. **Lingüística Computacional: teoria & prática**. São Paulo: Parábola, 2005.

SANTOS, L. L.; ARAGON, C. C.; GERARDI, F. Línguas minoritárias e anotações sintáticas de corpora: experiências de pesquisa na iniciação científica. **Letras de Hoje**, v. 59, n. 1, p. 1–9, 2024. DOI <https://doi.org/10.15448/1984-7726.2024.1.44734>

TORRENT, T. Plano brasileiro para turbinar IA ignora conceito básico da tecnologia. **Tilt Uol**, 23 jun. 2025. Disponível em: https://www.uol.com.br/tilt/analises/ultimas-noticias/2025/06/23/plano-brasileiro-para-turbinar-ia-ignora-conceito-basico-da-tecnologia.htm?utm_source=whatsapp-network&utm_medium=compartilhar_conteudo&utm_campaign=organica&utm_content=geral. Acesso em: 24 out. 2025.

UNIVERSITÄT TÜBINGEN. **Computational Linguistics**, 2025. Department of Linguistics, Studies & Teaching. Disponível em: <https://uni-tuebingen.de/fakultaeten/philosophische-fakultaet/fachbereiche/neuphilologie/seminar-fuer-sprachwissenschaft/studium-lehre/studiengaenge/computerlinguistik/>. Acesso em: 4 nov. 2025.

UNIVERSITY OF WASHINGTON. **Linguistics 567**, 2025. Linguistics 567: Knowledge Engineering for NLP. Disponível em: <https://courses.washington.edu/ling567/>. Acesso em: 4 nov. 2025.

UNIVERSITY OF WASHINGTON. **Languages Analyzed in Ling 567**, 2025. Disponível em: <https://wiki.ling.washington.edu/bin/view/Main/LanguagesList>. Acesso em: 4 nov. 2025.

ZARIQUIEY, R. *et al.* Building an Endangered Language Resource in the Classroom: Universal Dependencies for Kakataibo. In: CALZOLARI, N. *et al.* (org.). LREC 2022. **Proceedings of the Thirteenth Language Resources and Evaluation Conference**. Marseille, France: European Language Resources Association, jun. 2022. Disponível em: <https://aclanthology.org/2022.lrec-1.409>. Acesso em: 27 out. 2025.