Research Papers in Language Teaching and Learning
Vol. 16, No. 1, March 2026, 106-140
ISSN: 1792-1244
Available online at http://rpltl.eap.gr

Integrating Large Language Models into Corpus-Based Teaching: A Framework for Speech-Language Pathology, Computational Linguistics and Clinical Programs

Athanasios Karasimos, Evangelia-Antonia Efstratiadou, Christos Papatzalas & Ilias Papathanasiou

This paper proposes a systematic framework for integrating Large Language Models (LLMs) into corpus-based teaching for Speech-Language Pathology, Computational Linguistics, and Clinical Neurolinguistics programs. Traditional corpus analysis, though essential, presents significant pedagogical challenges, including being time-intensive and requiring specialized technical expertise, especially with atypical language data like aphasic speech. The emergence of LLMs offers a transformative opportunity to overcome these barriers. The systematic framework introduces practical educational modules, validated annotation workflows, and assessment strategies, exemplified through the Greek CACLA corpus. Crucially, the approach advocates for using LLMs as "cognitive partners" to handle routine tasks, allowing students to focus on higher-order analysis, clinical interpretation, and the critical evaluation of AI outputs. This method aims to democratize sophisticated linguistic analysis while ensuring students develop necessary critical thinking capacities and technological literacy.

Keywords: Large Language Models, Computational Linguistics, Speech-Language Pathology, Clinical Neurology, Aphasia, Corpus Methodology

---

1. Introduction

The integration of corpus linguistics into higher education has long been recognized as essential for training students in language sciences, such as speech-language pathology, computational linguistics, and clinical neuroscience programs (Reppen, 2010; Römer, 2011). However, traditional corpus analysis methods present significant pedagogical challenges: they are time-intensive, require specialized technical expertise, and often create barriers between linguistic theory and clinical application (Armstrong, 2000; Friginal & Hardy, 2014). Educators and students frequently struggle with the challenging learning curve of annotation software, and the difficulty of extracting meaningful patterns from clinical language data (Ball, 2013; Crystal, 2013). These challenges are particularly delicate when working with atypical language corpora, such as aphasic speech, where linguistic irregularities complicate automated annotations and analysis, and demand advanced interpretive skills (Goodglass *et al*., 2001).

The emergence of Large Language Models (LLMs), such as GPT-5, Gemini and Claude among others, represents a transformative opportunity for corpus-based education (Brown *et al*., 2020). These advanced AI systems demonstrate significant and evolving capabilities in linguistic annotation, pattern recognition, but still with several issues on natural language understanding (Bommasani *et al*., 2021; Zhao *et al*., 2025). Recent studies explore LLM applications in linguistic research, including morphological analysis (Mita *et al*., 2024; Karasimos & Makri, 2025; Karasimos, Makri & Petropoulou, 2025), syntactic parsing (Khan, 2025; Manning, 2022); yet their pedagogical integration remains largely unexplored, particularly in clinical and computational linguistics education or LLMs have been (extensively) mis-presented and ill-researched by non-AI-expertised educators.

Despite growing theoretical and research interest in AI-enhanced education (Holmes *et al*., 2019; Zawacki-Richter *et al*., 2019) and recognition of technology's role in language teaching and speech-language pathology training (Keck & Doarn, 2014; Hill *et al*., 2020), a significant research gap still remains; no comprehensive framework currently guides educators in integrating LLMs, corpora and educational technology tools into their curricula. This paper addresses this gap by presenting a systematic framework for incorporating LLMs and corpus-based teaching across speech-language pathology, computational linguistics, and clinical neurolinguistics courses. We introduce practical educational modules, validated annotation workflows, and assessment strategies, exemplified through the CACLA (Corpus for Aphasic Clinical Language Analysis) Greek corpus. Our aim is to provide educators, teachers, and practitioners with tried-and-tested tools that embed LLM capabilities while maintaining pedagogical aspect, critical thinking development, and awareness of technological limitations.

2. From Corpora to Classrooms: Bridging Clinical Language Data and Educational Practice

The path from analyzing aphasic speech corpora to developing effective educational frameworks requires understanding both the clinical linguistic landscape and the pedagogical challenges inherent in teaching corpus analysis methods. This section traces the evolution from traditional corpus analysis in communication disorders to the emerging potential of Large Language Models, while situating these developments within the broader context of technology-enhanced education in speech-language pathology and computational linguistics.

2.1 Corpus Analysis in Communication Disorders

Corpus-based approaches to studying communication disorders have transformed the understanding of atypical language patterns over the past three decades (MacWhinney *et al*., 2011; Ball *et al*., 2008). The systematic collection and analysis of language data from individuals with aphasia, language disorders, and other communication impairments offered the opportunity to identify quantifiable linguistic markers, patterns of recovery trajectories, and validate assessment protocols/tests (Armstrong, 2000; Berndt *et al*., 2000; Saffran *et al*., 1989; Wilson *et al*., 2017).

Clinical corpora serve multiple critical functions mostly in research (and less in education). They provide authentic data that capture the variability and complexity of disordered language production

(Prins & Bastiaanse, 2004; Stark *et al*., 2025; Varlokosta *et al*., 2016; Webster *et al*., 2007); for instance, the AphasiaBank corpus[1], contains hundreds of discourse samples from people with aphasia across diverse tasks and several languages, providing access to several phenomena ranging from word retrieval difficulties, error samples to discourse coherence (Dalton & Richardson, 2019; MacWhinney *et al*., 2011). Similarly, specialized corpora like the Northwestern Narrative Language Analysis (NNLA)[2] system provide structured protocols for analyzing story retelling, and offering insights into significant content and linguistic efficiency (Thompson *et al*., 2012; Kintz & Wright, 2018).

Based on the Greek digital landscape, corpus resources for communication disorders remain relatively limited, though several significant initiatives have emerged, such THALIS Aphasia (Varlokosta *et al*., 2016; Βαρλοκώστα *et al*., 2017) and some smaller individual research projects (Kambanaros & van Steenbrugge, 2006; Nanousi *et al*., 2006). The CACLA corpus represents a comprehensive effort to document Greek spoken and aphasic speech across various aphasia types and task conditions, providing rich material for both clinical research and educational applications (Papathanasiou *et al*., 2013).

The linguistic analysis of aphasic corpora presents unique challenges that distinguish it from typical language corpus research. Atypical language production includes phenomena such as *agrammatism* (simplified syntax with function word omissions), *paragrammatism* (syntactically complex but semantically empty output), *phonemic and semantic paraphasias* (sound-based and meaning-based word substitutions), *neologisms* (invented non-words or non-sense words), and discourse-level disruptions including tangentiality and reduced coherence (Caplan, 1987; Goodglass *et al*., 2001; Kolk, 1995). These features require experienced annotators to make complex linguistic judgments, distinguishing between intentional linguistic choices and pathological productions—a distinction that challenges both human and automated annotation systems (Rochon *et al*., 2000; Bastiaanse & Edwards, 2004).

Educational applications of clinical corpora have primarily focused on introducing students to authentic clinical data, teaching linguistic analysis skills, and illustrating theoretical concepts with real-world examples (Barbieri & Eckhardt, 2007; Römer, 2011). However, the labor-intensive nature of corpus annotation and analysis has limited widespread adoption in educational settings, particularly in programs lacking computational linguistics infrastructure (Friginal & Hardy, 2014; Reppen, 2010).

2.2 Large Language Models for (Spoken Data) Text Analysis

Recent research has begun exploring LLM applications across linguistic domains. Morphological analysis studies demonstrate high accuracy for part-of-speech tagging and morphological feature extraction across diverse languages, including morphologically complex ones (Liu *et al*., 2024; Üstün *et al*., 2024). For Greek specifically, LLMs show promise in handling rich inflectional morphology, though systematic evaluation remains limited (Koutsikakis *et al*., 2020).

Syntactic parsing evaluations reveal that LLMs can produce dependency and constituency parses approaching state-of-the-art specialized parsers, with particular strength in handling ambiguous structures through contextual reasoning (Kulmizev *et al*., 2024; Warstadt & Bowman, 2022). However, performance varies with sentence complexity and grammaticality—a critical consideration for clinical corpus analysis (Hu *et al*., 2020).

[1] https://talkbank.org/aphasia/
[2] https://talkbank.org/aphasia/discourse/C-NNLA/

Semantic analysis applications include word sense disambiguation, semantic role labeling, and metaphor detection, with LLMs often surpassing traditional systems by leveraging vast world knowledge encoded during pre-training (Pilehvar & Camacho-Collados, 2019; Ettinger, 2020). Discourse analysis capabilities extend to coreference resolution, coherence evaluation, and rhetorical structure identification, though systematic evaluation frameworks remain under development (Hao *et al*., 2021).

Error detection and correction represent particularly relevant applications for clinical linguistics. Studies show LLMs can identify grammatical errors, semantic anomalies, and stylistic inconsistencies, suggesting potential for automated coding of aphasic errors (Shen *et al*., 2023; Wu *et al*., 2023). However, the ability to distinguish pathological from intentional non-standard forms—crucial for clinical analysis—requires careful prompt engineering and validation (Bryant *et al*., 2023).

Cross-linguistic applications demonstrate that multilingual LLMs like GPT-5 and mBERT can perform reasonably well on lower-resourced languages, though performance typically degrades for languages with limited training data representation (Joshi *et al*., 2020; Üstün *et al*., 2024). For Greek, intermediate resource availability suggests moderate but not optimal LLM performance without fine-tuning (*Koutsikakis* et al., 2020).

The application of LLMs to atypical or clinical language data represents a relatively unexplored frontier. Preliminary investigations suggest both promise and challenges. LLMs demonstrate robustness to disfluency, maintaining reasonable parsing accuracy even with filled pauses, false starts, and self-corrections common in aphasic speech (Bhat *et al*., 2022). However, handling of agrammatism—telegraphic speech with systematic function word omissions—proves more challenging, as LLMs may inappropriately "correct" such productions rather than analyzing them as-produced.

Paraphasia detection studies (Zhu *et al*., 2023) suggest LLMs can identify semantic substitutions when provided with appropriate context and task framing, though phonemic paraphasias may require explicit phonological representation. Neologism identification presents particular difficulty, as LLMs may misinterpret non-words as rare real words, code-switching or hapax legomena, requiring explicit instruction to flag potential neologisms (Kuzman *et al*., 2023). Discourse-level analysis of clinical narratives shows promise for automated coherence rating and information content scoring. Studies applying GPT-based models to aphasia discourse samples report moderate correlations with human ratings of narrative quality (Fergadiotis *et al*., 2023). However, LLMs may over-emphasize semantic coherence while missing subtle linguistic markers of impairment, necessitating careful validation against clinical gold standards. Variability and consistency concerns emerge prominently in clinical applications. LLMs can produce different outputs for identical inputs across sampling runs, problematic for research requiring precise replicability (Ouyang *et al*., 2022). Temperature settings, random seeds, and prompt variations all influence output consistency—factors requiring systematic control in educational and research contexts (Sclar *et al*., 2023).

2.3 Technology in SLP Education

The integration of technology into speech-language pathology education has evolved substantially over recent decades, though adoption patterns vary considerably across institutions and geographic regions (Hill *et al*., 2020). Simulation and virtual patients represent well-established technologies in SLP education, allowing students to practice clinical interactions, assessment procedures, and intervention planning in risk-free environments (Dudding et al., 2011; Hill & Theodoros, 2017).

Platforms like SimuCase[3] provide standardized virtual cases covering diverse communication disorders, enabling repetitive practice and immediate feedback. However, these systems typically present scripted scenarios rather than authentic language samples, limiting exposure to real-world linguistic variability.

Video analysis tools enable students to review authentic clinical interactions, annotate communicative behaviors, and analyze language samples from recorded sessions. Software like ELAN[4] allows time-aligned transcription and annotation, though the labor-intensive nature limits typical use to small sample sizes (Sloetjes & Wittenburg, 2008). Computer-assisted learning modules deliver instructional content on anatomical structures, phonetic transcription, and assessment procedures through interactive multimedia (Brackenbury et al., 2008). While valuable for knowledge transmission, these typically lack integration with authentic clinical data analysis. Telepractice platforms facilitate remote clinical supervision, enabling students to observe real therapy sessions and receive feedback from clinical educators regardless of geographic location (Grogan-Johnson *et al*., 2013; Tucker, 2012). The COVID-19 pandemic accelerated adoption of telehealth modalities, demonstrating feasibility while highlighting technological barriers and training needs.

Specialized software for language analysis, including systematic analysis of language transcripts (SALT) and CLAN, provides quantitative analysis capabilities for clinical language samples (MacWhinney, 2000; Miller & Iglesias, 2020). However, educational surveys indicate limited integration into graduate curricula, attributed to perceived complexity, time investment, and insufficient instructor familiarity (Brackenbury *et al*., 2008). Broader healthcare education has increasingly incorporated artificial intelligence applications, providing relevant models for SLP education. Medical image analysis using deep learning algorithms helps radiology students develop diagnostic skills through immediate feedback on interpretation accuracy (Hosny *et al*., 2018; Shen *et al*., 2017). Clinical decision support systems powered by machine learning assist medical students in differential diagnosis, treatment planning, and predicting patient outcomes (Beam & Kohane, 2018; Rajkomar *et al*., 2019). Natural language processing applications in medical education include automated analysis of clinical notes, extraction of relevant information from electronic health records, and evaluation of documentation quality (Koleck *et al*., 2019; Wang *et al*., 2018). These applications demonstrate feasibility of integrating sophisticated AI tools into clinical training while maintaining pedagogical rigor. Intelligent tutoring systems adapt instructional content and feedback based on individual student performance, demonstrating learning gains compared to traditional instruction across medical disciplines (Kulik & Fletcher, 2016; Woolf *et al*., 2013). However, concerns regarding over-reliance on automation, reduced critical thinking, and "black box" decision-making have prompted calls for careful pedagogical design emphasizing human oversight and transparent reasoning (Holmes *et al*., 2019; Zawacki-Richter *et al*., 2019).

3. The educational prism of AI, LLMs and corpora in the tertiary system

The integration of artificial intelligence, Large Language Models, and corpus linguistics into tertiary education represents a convergence of technological innovation and pedagogical necessity. This section examines the theoretical foundations that underpin effective educational implementation, drawing from established learning frameworks, technology integration models, and corpus-based pedagogical research. We focus specifically on how these theoretical perspectives apply to computational linguistics, neurolinguistics, and speech-language pathology education—disciplines where authentic language data, analytical rigor, and clinical competence intersect.

[3] https://www.simucase.com/

[4] https://archive.mpi.nl/tla/elan

3.1 Educational Framework: Learning Objectives and Cognitive Development

The development of effective learning objectives for LLM-based corpus analysis requires careful alignment with program-specific competencies while addressing the unique cognitive demands of working with both clinical language data and artificial intelligence tools. Learning objectives must balance technical skill acquisition with critical evaluation capabilities, recognizing that students need not only to operate AI systems but also to understand their limitations and validate their outputs (Anderson & Krathwohl, 2001; Bloom et al., 1956; Fink, 2013).

For computational linguistics programs, learning objectives emphasize algorithmic understanding, evaluation methodology, and the ability to critically assess automated linguistic analysis. Students should be able to design annotation schemes, implement validation procedures, calculate inter-annotator agreement metrics, and compare LLM performance against traditional NLP approaches (Liddy, 2001; Jurafsky & Martin, 2023). These objectives align with computational thinking frameworks that emphasize decomposition of complex problems, pattern recognition, abstraction, and algorithmic design (Weintrop *et al*., 2016; Wing, 2006).

For speech-language pathology programs, learning objectives center on clinical application: identifying linguistically-based assessment measures, interpreting quantitative indices of language impairment, connecting corpus patterns to theoretical models of aphasia, and making evidence-based clinical decisions informed by linguistic analysis (American Speech-Language-Hearing Association [ASHA], 2016). The integration of corpus analysis with clinical reasoning represents what Fink (2013) terms "integration learning"—the ability to connect ideas, perspectives, and realms of life, essential for translating linguistic research into clinical practice.

For neurolinguistics and clinical neuroscience programs, objectives emphasize understanding brain-language relationships through systematic analysis of language breakdown patterns. Students must develop competencies in identifying neuroanatomical correlates of linguistic deficits, recognizing dissociations between preserved and impaired language functions, and critically evaluating neurolinguistic theories based on corpus evidence (Hillis, 2007; Papathanasiou *et al*., 2017).

3.2 Technology Integration in Teaching and Learning: Frameworks and Models TPACK Framework for AI-Enhanced Corpus Linguistics Education

The TPACK framework (Mishra & Koehler, 2006) guides integration of LLM-based corpus analysis into linguistics education, recognizing that effective technology use requires interplay between Technological (TK), Pedagogical (PK), and Content Knowledge (CK). TK encompasses LLM capabilities, prompt engineering, and computational principles underlying language models. PK involves scaffolding, assessment, and fostering critical thinking, while CK includes linguistic theory, clinical frameworks, and corpus methodology. The intersections are crucial, since TCK means understanding how AI tools address specific linguistic phenomena; PCK involves effective teaching strategies for complex concepts; TPK recognizes how LLMs enable students to analyze larger datasets, revealing patterns previously accessible only to researchers. Full TPACK integration—e.g., teaching agrammatism through collaborative LLM-assisted annotation that students validate—transforms not just what students learn but how (Koehler & Mishra, 2009).

The SAMR model (Puentedura, 2006) evaluates transformative potential across four levels: Substitution (LLMs replace manual annotation with no pedagogical change), Augmentation (students annotate larger datasets with immediate feedback), Modification (students compare LLM performance across aphasia types, integrating linguistic and computational thinking), and Redefinition (real-time analysis of large clinical corpora to test neurolinguistic hypotheses). Effective integration should target at least Modification level. From a constructivist perspective, LLM-based corpus analysis supports authentic learning through real clinical data, scaffolded learning via LLM-generated annotations students refine, and collaborative learning through group discussion of AI outputs. Communities of practice theory (Lave & Wenger, 1991) frames this as legitimate peripheral participation in professional linguistics communities.

3.3 Corpus Linguistics in Language-Related Education: Pedagogical Foundations and Innovations

Corpus linguistics has established significant pedagogical value across language-related disciplines, with data-driven learning (DDL) positioning students as researchers who discover linguistic patterns through direct corpus investigation rather than receiving pre-digested descriptions. This approach promotes inductive reasoning, pattern recognition, and empirical validation—cognitive skills essential for linguistic research and clinical practice (Boulton, 2012; Johns, 1991; 1998). Large language models enhance DDL by rapidly generating corpus patterns that students can examine, enabling focus on interpretation rather than data extraction.

Central to corpus-based education is the development of frequency and typicality awareness, as corpus analysis reveals which linguistic forms are common versus rare, typical versus marked. For clinical populations, corpus-based frequency norms inform assessment interpretation by determining whether a particular production pattern falls within or outside typical ranges for specific aphasia types (Berndt *et al*., 2000; Thompson et al., 2012). Students develop empirical grounding for clinical judgments through sustained corpus exposure.

Authentic language exposure distinguishes corpus-based education from constructed examples or idealized grammatical descriptions. Authentic clinical corpora reveal the complexity of real language production including disfluencies, self-corrections, and false starts, thereby preparing students for clinical realities (Gilquin & Granger, 2010). Exposure to variability within and across speakers develops realistic expectations and flexible analytical skills essential for professional practice. The investigative nature of corpus pedagogy encourages hands-on exploration where students actively query corpora, formulate hypotheses, test predictions, and refine analyses based on evidence. This investigative approach develops research literacy alongside linguistic knowledge, fostering skills transferable to professional contexts (Charles, 2015; Römer, 2011).

AI integration addresses many traditional barriers while creating new pedagogical possibilities. Reduced technical barriers emerge as natural language prompts replace complex query languages, making corpus analysis accessible to students without programming backgrounds. This accessibility enables accelerated analysis where students can examine multiple samples, compare across speakers or conditions, and iterate rapidly, transforming corpus analysis from isolated demonstration to integrated investigative method. Enhanced focus on interpretation occurs when annotation automation shifts student cognitive resources from mechanical coding to linguistic analysis and clinical reasoning. Students spend more time asking what patterns reveal about language processing rather than how to code particular constructions, aligning with constructivist principles emphasizing deep understanding over procedural execution. Integrated validation activities become pedagogically central as students compare LLM outputs against gold standards, identify systematic errors, and investigate why particular phenomena challenge automation. These metacognitive activities develop critical evaluation skills generalizable beyond corpus linguistics to any AI-enhanced professional context, while multilingual accessibility potentially expands through LLMs that handle diverse languages more flexibly than language-specific tools.

4. Design and Implementation Context

This paper presents a design-based framework developed and piloted across two Greek university programs during 2024–2025. The following sections describe the implementation context, corpus materials, and pedagogical design rationale. Full empirical evaluation of learning outcomes is ongoing and will be reported in subsequent work.

4.1 Research questions and aim

The overarching aim is to produce an evidence-based framework for LLM integration that balances technological innovation with pedagogical soundness while addressing discipline-specific needs.
RQ1: *To what extent can Large Language Models accurately perform multi-level linguistic annotation of Greek aphasic speech to support teaching scenarios and materials?*
RQ2: *Does integration of LLM-based corpus analysis enhance student learning outcomes compared to traditional corpus methods or no corpus instruction?*
RQ3: *What technical, pedagogical, and institutional barriers emerge during implementation, and what solutions prove effective?*
RQ4: *How do students perceive LLM-enhanced corpus analysis in terms of usability, learning value, and preparation for professional practice?*

4.2 Corpora Selection

The Greek CACLA (Corpus for Aphasic Clinical Language Analysis) serves as the primary corpus, comprising approximately 200 speech samples from Greek-speaking individuals with aphasia representing diverse types and severity levels (Papathanasiou *et al*., 2017). We curated a pedagogical subset of 36 transcripts (12 for demonstrations, 24 for student projects) selected for: (1) appropriate length (150-400 words); (2) linguistic representativeness of different aphasia types; (3) transcription quality following CHAT conventions; (4) task diversity; and (5) complexity gradient enabling differentiated instruction.

Gold standard annotations were developed through multi-stage expert consensus, with two linguists independently annotating each transcript following Universal Dependencies guidelines for Greek (Prokopidis & Papageorgiou, 2017) and clinical error coding systems (Schwartz *et al*., 2009). Disagreements (approximately 12%) were resolved through adjudication.

To accommodate different pedagogical objectives across programs, we incorporated comparison datasets enabling contrastive analysis. For English/Linguistics students, we included a parallel corpus of 10 transcripts from neurologically intact Greek speakers performing identical elicitation tasks (picture description, narrative retelling, procedural discourse). This typical language corpus enables students to systematically compare linguistic patterns between aphasic and normal speech—identifying which features represent pathological deviation versus normal variability, examining quantitative differences in productivity and complexity measures, and developing empirically-grounded understanding of linguistic breakdown patterns (Armstrong, 2000; Goodglass *et al*., 2001). This contrastive approach aligns with computational linguistics pedagogy emphasizing comparative analysis and data-driven pattern recognition (Römer, 2011). Additionally, we incorporated AphasiaBank English samples for cross-linguistic comparison (MacWhinney et al., 2011), and outputs from traditional NLP tools (Greek spaCy, Stanza) for comparative evaluation of LLM versus rule-based/statistical annotation approaches—particularly relevant for students focusing on computational methods.

4.3 Study design behind teaching procedures

The study employs a quasi-experimental pre-post design with comparison groups implemented across two Greek university departments during 2024-2025. The first group (n ≈ 55) receives a full LLM-enhanced curriculum including prompt engineering, validation methodology, and collaborative analysis projects. The second group (n ≈ 67) receives AI-modified clinical training with systematic corpus analysis and aphasic speech data in two different modules.

All groups complete identical pre/post tests assessing corpus linguistics knowledge, clinical language understanding, and analytical skills. Data collection employs convergent mixed-methods design (Creswell & Plano Clark, 2017): quantitative measures (knowledge tests, annotation accuracy, Likert-scale surveys) and qualitative data (focus groups, observation checklists). Statistical analysis includes mixed-model ANOVAs examining pre-post gains across groups, t-tests comparing practical competency, and regression models predicting outcomes from background variables. Qualitative data undergoes thematic analysis following Braun and Clarke (2006).

4.4 Participants

Participants comprise 122 third- and fourth-year undergraduate students across two programs: School of English (n = 55) and Speech-Language Pathology (n = 67). The sample is predominantly female (85%), aged 20-26 years (M = 21.8), consistent with linguistics and SLP program demographics internationally. All are native Greek speakers or have advanced proficiency (C1/C2; 9 bilinguals with Russian, Albanian, Ukrainian and Turkish). Prior coursework includes foundational linguistics, though only 23% report corpus linguistics exposure and 12% have programming experience. Technology experience is mixed: 67% have used ChatGPT (or any other LLM) for general purposes, but only 15% for academic analytical tasks. This profile suggests students can navigate AI interfaces but lack critical evaluation skills for linguistic applications—a key learning objective.

5. Educational Applications

5.1 Curriculum Integration

The LLM-based corpus analysis framework integrates across three distinct but complementary curricular modules, each addressing specific disciplinary competencies while leveraging shared technological infrastructure. The Theoretical Module (*Aphasia and Related Disorders*) emphasizes neurocognitive models of language breakdown, classification and symptomatology of acquired aphasia and related disorders, and evidence-based principles of language assessment and rehabilitation—developing foundational theoretical competence in adult neurogenic communication disorders. The Digital Module (*Computational Linguistics and Natural Language Processing*) emphasizes annotation methodology, prompt engineering for linguistic tasks, and systematic evaluation of automated analysis tools—developing technical competencies in NLP applications. The Clinical Module (*Clinical Practice III*) emphasizes supervised application of assessment, differential diagnosis, intervention planning, and professional documentation in real clinical settings—developing advanced clinical reasoning, ethical decision-making, and entry-level professional competence in adult speech-language pathology. This module enables students to consolidate theoretical knowledge and digital competencies into advanced, context-sensitive clinical decision-making. This three-pronged approach ensures students develop theoretical understanding, technical proficiency, and clinical reasoning—essential competencies for contemporary speech-language pathology and computational linguistics professionals (ASHA, 2016; Jurafsky & Martin, 2023).

5.2 Specific Educational Modules

5.2.1 Theoretical Module: Aphasia and Related Disorders

The Theoretical Module Aphasia and Related Disorders provides the conceptual and neurocognitive foundation for the integrated curriculum and is implemented as a structured lecture sequence (9 instructional hours; for more details see Appendix A) designed to develop theoretical coherence, analytic depth, and preparedness for advanced clinical application. The module addresses a longstanding challenge in speech-language pathology education: enabling students to move beyond surface-level classification toward principled understanding of how neurological damage disrupts language systems across modalities and levels of linguistic organization (Goodglass & Wingfield, 1997; Caplan & Marshall,1992). In Session 1, Foundations of Aphasia and Neurogenic Language Breakdown, students are introduced to historical and contemporary definitions of aphasia, etiological factors, and neuroanatomical correlates of language processing. Classical aphasia syndromes are presented alongside modern neurocognitive and functional models, highlighting both their heuristic value and their limitations in capturing real-world communicative functioning (Bastiaanse & Prins, 2013; Dronkers et al., 2017). Through analysis of authentic aphasic language samples, students begin to recognize inter- and intra-speaker variability and to critically evaluate the adequacy of syndromic labels. This session supports a shift from taxonomic memorization to explanatory, model-based reasoning about language breakdown. Session 2, Language Assessment and Breakdown Patterns, focuses on systematic evaluation of spoken and written language across comprehension, expression, naming, repetition, reading, writing, and discourse. Students examine standardized and non-standardized assessment approaches and learn to interpret error patterns in relation to underlying linguistic and cognitive processes, such as lexical access, morphosyntactic encoding, and discourse planning (Chapey, 2001). Particular emphasis is placed on distinguishing aphasia from co-occurring motor speech and cognitive-communication disorders, reinforcing diagnostic precision and reducing misclassification in clinical contexts. In Session 3, Aphasia, Related Disorders, and Functional Communication, the focus expands to acquired language disorders associated with traumatic brain injury and dementia. Students explore how diffuse versus focal neuropathology affects language, cognition, and communicative participation, and how progressive versus non-progressive conditions shape assessment and intervention priorities (Chapey, 2001; Bayles & Tomoeda, 2014). The session cites aphasia within the WHO–ICF framework, emphasizing activity limitations and participation restrictions alongside impairment-level analysis (WHO, 2001). Conceptual principles of intervention are introduced, highlighting evidence-based practice, functional goal setting, and the role of communication partners (Simmons-Mackie et al., 2010). Pedagogically, this module prioritizes conceptual integration and theoretical grounding over procedural skill acquisition. By developing explanatory models of language breakdown, it provides the cognitive scaffolding required for both computational analysis in the Digital Module and applied clinical reasoning in Clinical Practice III, supporting deep learning and transfer across contexts (National Research Council,, 2000).

5.2.2 Digital Module: Computational Linguistics and Natural Language Processing

The Digital Module represents the core computational linguistics application of LLM-based corpus analysis, implemented as a four-session sequence (5 instructional hours; for more details see Appendix A) designed to develop both technical proficiency and critical evaluation skills. This module addresses a critical gap in linguistics education: making sophisticated corpus annotation accessible to students without programming backgrounds while maintaining methodological straightforwardness. In session 1, Foundations and Initial Exploration introduces students to the CACLA corpus structure, Greek morphosyntactic complexity, and traditional annotation challenges. Students conduct manual annotation of a brief sample (15-20 words), documenting time investment and difficulties encountered. This experiential foundation establishes appreciation for automation benefits while grounding students in linguistic fundamentals—preventing what Pea (2004) terms "cognitive offloading" where technology bypasses essential conceptual development. The session concludes with a live demonstration of Claude performing identical annotation tasks in seconds, creating productive cognitive dissonance that motivates deeper investigation of how LLMs accomplish linguistic analysis.

During session 2, Prompt Engineering and Annotation Generation shifts to hands-on LLM interaction. Working in small groups, students design prompts for morphological annotation (POS tagging, lemmatization, feature extraction) using provided templates as starting points. The iterative prompt refinement process—testing, evaluating output quality, modifying instructions, retesting—develops metacognitive awareness about linguistic specification and computational interpretation (White *et al*., 2023). Students discover that effective prompts require explicit articulation of annotation conventions, output formats, and edge case handling—essentially requiring them to formalize their linguistic knowledge computationally. This "teaching the machine" process deepens linguistic understanding while developing prompt engineering as an emerging professional skill (Kasneci *et al*., 2023).

In the 3rd session, Validation and Critical Evaluation represents the pedagogically crucial component distinguishing our approach from uncritical AI adoption. Students systematically compare their LLM-generated annotations against expert gold standards, calculating precision, recall, and F1-scores for different annotation types. Error analysis reveals systematic patterns: LLMs excel at standard morphological forms but struggle with neologisms, phonemic paraphasias, and agrammatic structures lacking function words (Kuzman *et al*., 2023). Students categorize errors, hypothesize about computational causes (training data bias, context limitations, lack of clinical linguistic knowledge), and discuss implications for research applications. This critical engagement prevents "automation bias"—the tendency to over-trust automated outputs—while developing an understanding of AI capabilities and constraints (Goddard *et al*., 2012).

Finally, during session 4, Comparative Analysis and Synthesis culminates in contrastive examination of aphasic versus typical language corpora. Students use their validated LLM workflows to analyze multiple samples, extracting quantitative measures (MLU, TTR, grammatical accuracy rates) and identifying qualitative patterns. The larger dataset sizes enabled by automation allow observation of cross-speaker generalizations impossible with manual methods within course timeframes. Final presentations synthesize technical findings with linguistic interpretation: What do corpus patterns reveal about Greek morphological processing? How do computational analysis results align with or challenge theoretical models of agrammatism? This integration of technical and theoretical dimensions exemplifies Bloom's highest taxonomic levels—synthesis and evaluation (Anderson & Krathwohl, 2001). Pedagogical significance lies in balancing automation efficiency with analytical depth, technical skill development with critical thinking, and computational methods with linguistic insight—preparing students for futures where AI tools are ubiquitous but human expertise remains essential.

5.2.3 Clinical Module: Clinical Practice III

The Clinical Module *Clinical Practice III* represents the culminating application of theoretical and analytical competencies within supervised real-world clinical environments. Implemented as a semester-long clinical placement with supporting lectures (10 ECTS), the module is designed to support students' transition from guided learning to entry-level professional practice. Its central pedagogical objective is the development of integrated clinical reasoning, ethical judgment, and professional autonomy under supervision, consistent with international standards for clinical education in speech-language pathology (ASHA, 2016; RCSLT, 2017). In Session 1, Advanced Clinical

Assessment and Case Formulation, students engage in comprehensive evaluation of adults with acquired communication disorders, including structured case history intake, orofacial and cranial nerve examination, cognitive screening, and detailed language assessment. Drawing on theoretical knowledge from the *Aphasia and Related Disorders* module, students synthesize linguistic, cognitive, and neurological data into coherent clinical hypotheses. Emphasis is placed on differential diagnosis in complex cases involving overlapping aphasic, cognitive-communication, and motor speech features, reflecting best practices in adult neurogenic assessment (Papathanasiou, Coppens, & Potagas, 2017). Session 2, Clinical Decision-Making and Intervention Planning, focuses on translating assessment findings into evidence-based therapy goals and intervention plans. Students are trained to prioritize functional communication outcomes, align goals with patient needs and contextual factors, and justify clinical decisions using empirical evidence and theoretical rationale (Chapey, 2001). Supporting lectures on acquired language disorders in traumatic brain injury and dementia reinforce the importance of flexible clinical reasoning and longitudinal planning in progressive and non-progressive conditions. In Session 3, Documentation, Outcome Monitoring, and Professional Practice, students develop advanced skills in clinical documentation, including evaluation reports, therapy plans, progress notes, and SOAP documentation. Cognitive and quality-of-life questionnaires are incorporated into outcome monitoring, reinforcing a holistic and patient-centered approach to clinical effectiveness (Hilari & Byng, 2009). Ethical considerations—including informed consent, confidentiality, professional accountability, and interdisciplinary collaboration—are embedded throughout clinical practice, aligning with professional codes of ethics and legal frameworks (ASHA, 2016). Pedagogically, Clinical Practice III exemplifies experiential learning at the highest level, integrating theoretical knowledge, analytical competence, and procedural skill within authentic clinical contexts. By the end of the module, students demonstrate readiness for professional practice, characterized by evidence-based clinical reasoning, ethical awareness, reflective capacity, and adaptability to complex real-world communication disorders.

5.3 Towards a common educational scenario template

The proposed educational scenario template provides a flexible, adaptable framework for integrating LLMs and corpus analysis across diverse linguistic and clinical contexts beyond the specific CACLA implementation. This modular template follows a four-phase pedagogical progression that can be customized for different languages, corpora, and target competencies: Phase 1 (Foundation and Manual/Traditional Exploration) establishes baseline understanding through hands-on engagement with the chosen corpus and traditional annotation methods; Phase 2 (AI-Enhanced Pipeline Development) introduces LLM integration through systematic prompt engineering and automation workflows; Phase 3 (Validation and Critical Analysis) develops evaluation competencies through systematic comparison with gold standards and error analysis; and Phase 4 (Synthesis and Application) culminates in comprehensive analysis and presentation of findings. The template's strength lies in its adaptability—instructors can substitute alternative corpora (clinical, literary, social media, multilingual), modify linguistic annotation levels (phonological, morphological, syntactic, discourse), adjust technical complexity for different student populations, and customize assessment criteria for specific learning objectives. Essential template components include prerequisite specification, resource requirements, step-by-step process documentation, variation suggestions for different contexts, and assessment rubrics, enabling educators to create robust LLM-enhanced corpus linguistics curricula tailored to their institutional needs and student populations while maintaining pedagogical rigor and critical thinking development.

6. Discussion

6.1 Pedagogical Rationale and Anticipated Outcomes

The Theoretical Module was designed to foster foundational competencies that transformed their understanding of neurological language disorders from superficial classification to principled theoretical reasoning. Through systematic engagement with Session 1 (Foundations of Aphasia and Neurogenic Language Breakdown), students will learn to critically evaluate classical aphasia syndromes alongside modern neurocognitive models, recognizing both their heuristic value and limitations in capturing real-world communicative functioning. The analysis of authentic aphasic language samples will enable students to appreciate inter- and intra-speaker variability while moving beyond taxonomic memorization toward explanatory, model-based reasoning about language breakdown. Session 2 (Language Assessment and Breakdown Patterns) will equip students with systematic evaluation skills across comprehension, expression, and discourse modalities, teaching them to interpret error patterns in relation to underlying linguistic and cognitive processes such as lexical access and morphosyntactic encoding. Most significantly, Session 3 (Aphasia, Related Disorders, and Functional Communication) will broaden students' clinical perspective to encompass acquired language disorders in traumatic brain injury and dementia, while situating aphasia within the WHO-ICF framework to emphasize activity limitations and participation restrictions alongside impairment-level analysis.

The Digital Module will successfully bridge the gap between theoretical linguistic knowledge and practical computational application, enabling students to develop both technical proficiency and critical evaluation expertise. Through Session 1 (Foundations and Initial Exploration), students will gain experiential understanding of annotation challenges by manually processing brief samples before witnessing LLM capabilities, creating productive cognitive dissonance that will motivate deeper investigation of computational linguistic analysis. Session 2 (Prompt Engineering and Annotation Generation) will teach students to articulate their linguistic knowledge computationally through iterative prompt refinement, discovering that effective prompts require explicit specification of annotation conventions, output formats, and edge case handling—essentially formalizing their linguistic understanding for computational interpretation. The pedagogically crucial Session 3 (Validation and Critical Evaluation) will develop systematic comparison skills as students calculated precision, recall, and F1-scores while conducting error analysis that will reveal LLM strengths in standard morphological forms and limitations with neologisms, phonemic paraphasias, and agrammatic structures. Session 4 (Comparative Analysis and Synthesis) will enable students to leverage automation for examining larger datasets, extracting quantitative measures like MLU and TTR while synthesizing technical findings with linguistic interpretation to address questions about Greek morphological processing and theoretical models of agrammatism.

Students in Clinical Practice III will achieve the critical transition from guided academic learning to entry-level professional practice through integration of theoretical knowledge, analytical competencies, and supervised clinical experience. Session 1 (Advanced Clinical Assessment and Case Formulation) will teach students to synthesize linguistic, cognitive, and neurological data into coherent clinical hypotheses while engaging in comprehensive evaluation including case history intake, orofacial examination, and detailed language assessment. The emphasis on differential diagnosis in complex cases involving overlapping aphasic, cognitive-communication, and motor speech features will prepare students for real-world clinical complexity. Session 2 (Clinical Decision-Making and Intervention Planning) developed evidence-based clinical reasoning as students will learn to translate assessment findings into functional therapy goals, prioritize communication outcomes, and justify clinical decisions using empirical evidence and theoretical rationale. Session 3 (Documentation, Outcome Monitoring, and Professional Practice) will equip students with advanced clinical documentation skills including evaluation reports and SOAP documentation, while embedding ethical considerations throughout clinical practice. By module completion, students will demonstrate readiness for professional practice characterized by evidence-based clinical reasoning, ethical awareness, reflective capacity, and adaptability to complex real-world communication disorders.

6.2 Educational Effectiveness

The advantages outlined in section 6.3 demonstrate LLMs' transformative potential for corpus-based education in linguistics and speech-language pathology. Primary benefits include democratized accessibility through elimination of programming prerequisites and natural language prompting that replaces complex query languages, enabling sophisticated corpus analysis across diverse technical backgrounds. Time efficiency gains facilitate qualitative pedagogical shifts from analyzing 2-3 brief samples to examining dozens of transcripts, revealing cross-speaker patterns essential for clinical competence while redirecting cognitive resources from mechanical coding toward higher-order linguistic analysis and clinical reasoning. For clinical applications, LLMs provide rapid extraction of quantitative language measures that inform evidence-based assessment and intervention outcomes, with particular value for under-resourced languages like Greek where specialized NLP tools remain limited. The pedagogical flexibility allows instructors to customize difficulty and scaffold complexity progressively, making LLM integration essential for interdisciplinary contexts spanning computational linguistics to clinical practice.

6.3 LLMs for Corpus Analysis, Language and Speech Therapy: Advantages

Large Language Models offer transformative advantages for corpus-based education in linguistics and speech-language pathology that address longstanding pedagogical barriers. Accessibility and democratization represent primary benefits: LLMs eliminate programming prerequisites, making sophisticated corpus analysis available to students across diverse technical backgrounds (Kasneci et al., 2023). Natural language prompting replaces complex query languages, lowering entry barriers that historically limited corpus linguistics to computationally-trained specialists (Reppen, 2010).

Time efficiency gains enable qualitative pedagogical shifts. Where manual annotation constrained students to analyzing 2-3 brief samples, LLM-assisted workflows permit examination of dozens of transcripts, revealing cross-speaker patterns and population-level generalizations essential for clinical and research competence (Boulton & Cobb, 2017). This scalability transforms corpus analysis from isolated demonstration to integrated investigative method. Enhanced focus on interpretation emerges as automation handles mechanical coding, redirecting cognitive resources toward linguistic analysis, clinical reasoning, and theoretical synthesis—higher-order thinking central to professional education (Anderson & Krathwohl, 2001; Fink, 2013).

For clinical applications, LLMs provide rapid extraction of quantitative language measures (productivity, complexity, accuracy indices) that inform evidence-based assessment, benchmark intervention outcomes, and support clinical documentation requirements increasingly demanded in healthcare contexts (ASHA, 2016; Thompson *et al*., 2012). Multilingual capabilities particularly benefit under-resourced languages like Greek, where specialized NLP tools remain limited; multilingual LLMs offer reasonable performance without extensive language-specific development (Joshi et al., 2020). Finally, pedagogical flexibility allows instructors to customize difficulty, scaffold complexity progressively, and adapt activities to diverse learning objectives—essential for interdisciplinary contexts spanning computational linguistics to clinical practice.

6.4 Limitations and Challenges

Despite significant advantages, LLM integration presents substantial limitations requiring careful pedagogical management. Technical limitations constrain reliability for critical applications. Annotation accuracy varies considerably by linguistic level and phenomenon: while morphological tagging achieves 85-92% accuracy for standard Greek forms, performance degrades substantially for clinical language features including neologisms (62% accuracy), phonemic paraphasias (71%), and agrammatic structures with systematic omissions (68%) (cf. Kuzman et al., 2023; Liu et al., 2024). Context understanding remains shallow; LLMs lack genuine comprehension of clinical presentations, neuroanatomical substrates, or theoretical frameworks guiding interpretation. Hallucination risks—confident generation of plausible but incorrect analyses—pose serious concerns when students lack expertise to recognize errors (Alkaissi & McFarlane, 2023).

Educational considerations demand vigilant attention. Over-reliance on automation without developing foundational skills represents the primary pedagogical risk; students may generate sophisticated-appearing analyses without grasping underlying linguistic principles—what Salomon et al. (1991) term "intelligence in the system rather than intelligence with the system." Critical thinking development requires deliberate scaffolding; validation exercises, error analysis, and comparative evaluation must be integrated systematically rather than optional supplements. Technical literacy requirements extend beyond operational skills to conceptual understanding of LLM capabilities, limitations, and appropriate use cases—a new form of literacy not yet systematically addressed in linguistics curricula (Cardon *et al*., 2023).

Practical barriers complicate implementation. Cost considerations vary: while free LLM tiers enable basic exploration, intensive educational use may require institutional subscriptions ($20-30/student/month), substantial in resource-constrained contexts. Privacy and data security concerns arise with clinical corpora containing sensitive information, even when de-identified; institutional policies may restrict uploading patient-derived data to commercial AI platforms. Instructor expertise gaps emerge as most linguistics faculty lack experience with prompt engineering, AI evaluation methodology, or pedagogical strategies for critical technology engagement—necessitating substantial professional development investment (Mishra & Koehler, 2006; Zawacki-Richter et al., 2019).

6.5 Best Practices for Implementation

Evidence from our implementation suggests several critical practices for successful LLM integration. Foundational skills first: Students must develop basic linguistic analysis competencies through manual annotation before automation introduction, ensuring conceptual understanding precedes technological efficiency (Pea, 2004). Explicit prompt engineering requries students to record, version, and justify all prompts develops metacognitive awareness about linguistic specification and creates reproducible workflows aligned with research integrity standards (White et al., 2023). Scaffolded complexity progression is eminent by beginning with straightforward annotation tasks (POS tagging) before advancing to challenging phenomena (discourse coherence, clinical error detection) builds confidence while revealing AI limitations organically. Finally, interdisciplinary collaboration will rise by pairing computational linguistics and clinical students leverages complementary expertise, models professional teamwork, and enriches learning through diverse perspectives.

6.6 Future Directions

Future development should address several critical directions. Longitudinal evaluation tracking whether corpus analysis skills and critical AI literacy persist beyond course completion and transfer to professional practice remains essential for validating educational impact (Kirkpatrick & Kirkpatrick, 2006). Cross-linguistic expansion to additional under-resourced languages could democratize corpus linguistics globally, though requiring systematic validation of LLM performance across diverse linguistic typologies (Joshi *et al*., 2020). Hybrid approaches integrating LLMs with specialized clinical NLP tools may optimize accuracy-efficiency trade-offs by using LLMs for initial annotation and traditional tools for validation, or vice versa (Bryant *et al*., 2023). Custom fine-tuning of open-source models on clinical language corpora could enhance performance on atypical language while addressing privacy concerns through local deployment, though requiring institutional computational infrastructure (Üstün *et al*., 2024).

To achieve curriculum standardization, it is needed to develop consensus guidelines for AI literacy competencies in linguistics and SLP programs would ensure systematic preparation for technology-integrated professional futures. Ethical framework development specifically addressing AI use with vulnerable populations, clinical data, and healthcare applications remains urgent as technology outpaces regulatory guidance (Mittelstadt *et al*., 2016). Finally, authentic clinical integration can be achieved by moving beyond educational simulation to implementing LLM-assisted corpus analysis in actual clinical assessment, treatment planning, and outcome documentation would demonstrate real-world value while identifying practical implementation barriers. Collaborative research between educational institutions, healthcare facilities, and AI developers could accelerate responsible clinical translation benefiting both professional training and patient care.

7. Conclusions

The three-module structure — Theoretical, Computational, and Clinical — provides a principled progression from foundational neurolinguistic knowledge through computational methodology to supervised clinical application, while the four-phase scenario template gives instructors in diverse institutional contexts a reusable structure adaptable to other languages, corpora, and learning objectives. Throughout, the framework positions LLMs as “digital co-educators” rather than authoritative tools: by requiring students to validate automated outputs against gold standards and to analyse systematic errors, it builds critical evaluation skills and technological literacy alongside disciplinary knowledge, counteracting the automation bias that uncritical AI adoption risks.

Several limitations should be acknowledged; full empirical evaluation of learning outcomes is ongoing, and the current implementation is confined to two Greek university programs; generalisability to other national and linguistic contexts remains to be established. Reliance on commercial LLM platforms also raises sustainability and data-privacy concerns when working with patient-derived clinical material, and successful implementation presupposes instructors with competence in both prompt engineering and clinical linguistics — expertise not yet widely available in linguistics faculties.

For educators, the framework offers a concrete, theoretically grounded entry point for curriculum development that requires no programming background in instructors or students. For the field more broadly, it models responsible AI integration in clinical professional education — treating technological literacy as inseparable from disciplinary expertise and ethical awareness. Priorities for future work include longitudinal evaluation of skill transfer to professional practice, cross-linguistic validation, and the development of open-source or locally deployable alternatives to commercial LLMs for use with sensitive clinical corpora.

Acknowledgments

The research project was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “Basic Research Financing (Horizontal Support for All Sciences), National Recovery and Resilience Plan (Greece 2.0)” (Project Number: 016344). The principal investigator was Ilias Papathanasiou.

References

American Speech-Language-Hearing Association [ASHA]. (2016). *Scope of practice in speech-language pathology*. Retrieved from https://www.asha.org/policy/

American Speech-Language-Hearing Association [ASHA]. (2023). *2023 membership and affiliation profile*. https://www.asha.org/

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Anil, R., Dai, A. M., Firat, O., … & Wu, Y. (2023). PaLM 2 technical report. *arXiv preprint*. https://doi.org/10.48550/arXiv.2305.10403

Anthropic. (2024). *Claude 3 model card*. Retrieved from https://www.anthropic.com

Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology*, 14(9), 875-892.

Ball, M. J. (2013). *Research methods in clinical linguistics and phonetics: A practical guide*. Wiley-Blackwell.

Ball, M. J., Perkins, M. R., Müller, N., & Howard, S. (Eds.). (2008). *The Handbook of Clinical Linguistics*. Blackwell Publishing.

Barbieri, F., & Eckhardt, S. (2007). Applying corpus-based findings to form-focused instruction: The case of reported speech. *Language Teaching Research*, 11(3), 319-346. https://doi.org/10.1177/136216880707756

Bardovi-Harlig, K., & Mossman, S. (2016). Corpus-based materials development for teaching and learning pragmatic routines. In B. Tomlinson (Ed.), *SLA research and materials development for language learning* (pp. 250-267). Routledge.

Bastiaanse, R., & Edwards, S. (2004). Word order and finiteness in Dutch and English Broca's and Wernicke's aphasia. *Brain and Language*, 89(1), 91-107. https://doi.org/10.1016/S0093-934X(03)00306-7

Bastiaanse, R., & van Zonneveld, R. (2005). Sentence production with verbs of alternating transitivity in agrammatic Broca's aphasia. *Journal of Neurolinguistics*, 18(1), 57-66. https://doi.org/10.1016/j.jneuroling.2004.11.006

Bastiaanse, R., & Prins, R. S. (2013). Aphasia. In L. Cummings (Ed.), *The Cambridge handbook of communication disorders* (pp. 224–246). Cambridge University Press.

Bayles, K. A., & Tomoeda, C. K. (2014). *Cognitive-communication disorders of dementia: Definition, diagnosis, and treatment* (2nd ed.). Plural Publishing Inc.

Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA*, 319(13), 1317-1318. https://doi.org/10.1001/jama.2017.18391

Belland, B. R., Kim, C., & Hannafin, M. J. (2013). A framework for designing scaffolds that improve motivation and cognition. *Educational Psychologist*, 48(4), 243-270. https://doi.org/10.1080/00461520.2013.838920

Berndt, R. S., Wayland, S., Rochon, E., Saffran, E., & Schwartz, M. (2000). *Quantitative production analysis*. Psychology Press.

Bhat, S., Culotta, A., & Bhamidipati, N. (2022). *DisfluencyFixer: A tool to enhance Language Learning through Speech To Speech Disfluency Correction*. arXiv preprint. arXiv:2305.16957. https://doi.org/10.48550/arXiv.2305.16957

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Bitzenbauer, P. (2023). ChatGPT in physics education: A pilot study on easy-to-implement activities. *Contemporary Educational Technology*, 15(3), ep430. https://doi.org/10.30935/cedtech/13176

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. David McKay.

Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* [Computer program]. Version 6.4.60.

Bommasani, R., Hudson, D. A., Adeli, E., et al. & Liang, P. (2021). *On the opportunities and risks of foundation models*. arXiv preprint arXiv:2108.07258. https://doi.org/10.48550/arXiv.2108.07258

Boulton, A. (2012). Corpus consultation for ESP: A review of empirical research. In A. Boulton, S. Carter-Thomas, & E. Rowley-Jolivet (Eds.), *Corpus-informed research and learning in ESP: Issues and applications* (pp. 261-291). John Benjamins. https://doi.org/10.1075/scl.52.11bou

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348-393. https://doi.org/10.1111/lang.12224

Brackenbury, T., Burroughs, E., & Hewitt, L. E. (2008). A qualitative examination of current guidelines for evidence-based practice in child language intervention. *Language, Speech, and Hearing Services in Schools*, 39(1), 78-88. https://doi.org/10.1044/0161-1461(2008/008)

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). National Academy Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., …D. Amodei. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901. https://doi.org/10.48550/arXiv.2005.14165

Bryant, C., Yuan, Z., Qorib, M. R., et al. (2023). Grammatical error correction: A survey of the state of the art. *Computational Linguistics*, 49(3), 1-59. https://doi.org/10.1162/coli_a_00478

Bubeck, S., Chandrasekaran, V., Eldan, R., …, Zhang, L. (2023). *Sparks of artificial general intelligence: Early experiments with GPT-4*. arXiv preprint arXiv:2303.12712. https://doi.org/10.48550/arXiv.2303.12712

Caplan, D. (1987). *Neurolinguistics and linguistic aphasiology: An introduction*. Cambridge University Press.

Caplan, D., & Marshall, J. C. (Eds.) (1992). *Language: Structure, processing, and disorders*. The MIT Press.

Cardon, P. W., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023). The challenges and opportunities of AI-assisted writing: Developing AI literacy for the AI age. *Business and Professional Communication Quarterly*, 86(3), 257-295. https://doi.org/10.1177/23294906231176

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1), 43. https://doi.org/10.1186/s41239-023-00411-8

Chapey, R. (2001). *Language intervention strategies in aphasia and related neurogenic communication disorders* (4th ed.). Lippincott Williams & Wilkins.

Charles, M. (2015). Same task, different corpus: The role of personal corpora in EAP classes. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 131-154). John Benjamins. https://doi.org/10.1075/scl.69.07cha

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type-token ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2), 94-100. https://doi.org/10.1080/09296171003643098

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Crosthwaite, P. (2021). *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. Routledge.

Crystal, D. (2013). *Profiling linguistic disability* (3rd ed.). Wiley-Blackwell.

Dalton, S. G., & Richardson, J. D. (2019). A large-scale comparison of main concept production between persons with aphasia and persons without brain injury. *American Journal of Speech-Language Pathology*, 28(1S), 293-320. https://doi.org/10.1044/2018_AJSLP-17-0166

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT* (pp. 4171-4186). https://doi.org/10.18653/v1/N19-1423

Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.*), Collaborative-learning: Cognitive and computational approaches* (pp. 1-19). Elsevier.

Dillenbourg, P., Järvelä, S., & Fischer, F. (2016). *The evolution of research on computer-supported collaborative learning*. Springer.

Dronkers, N. F., Ivanova, M. V., & Baldo, J. V. (2017). What do language disorders reveal about brain–language relationships? From classic models to network approaches. *Journal of the International Neuropsychological Society*, *23*(9–10), 741–754. https://doi.org/10.1017/S1355617717001126

Ettinger, A. (2020). What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. *Transactions of the Association for Computational Linguistics*, 8, 34-48. https://doi.org/10.48550/arXiv.1907.13528

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. Aphasiology, 25(11), 1414-1430. https://doi.org/10.1080/02687038.2011.603898

Fergadiotis, G., Wright, H. H., & Capilouto, G. J. (2011). Productive vocabulary across discourse types. *Aphasiology*, 25(10), 1261-1278. https://doi.org/10.1080/02687038.2011.606974

Fink, L. D. (2013). *Creating significant learning experiences: An integrated approach to designing college courses* (2nd ed.). Jossey-Bass.

Friginal, E., & Hardy, J. A. (Eds.). (2014). *Corpus-based sociolinguistics: A guide for students*. Routledge.

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 359-370). Routledge.

Goodglass, H. (1993). *Understanding aphasia*. Academic Press.

Goodglass, H., & Wingfield, A. (Eds.). (1997). *Anomia: Neuroanatomical and cognitive correlates*. Academic Press.

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *Boston diagnostic aphasia examination* (3rd ed.). Lippincott Williams & Wilkins.

Grogan-Johnson, S., Gabel, R. M., Taylor, J., Rowan, L. E., Alvares, R., & Schenker, J. (2013). A pilot investigation of speech sound disorder intervention delivered by telehealth to school-age children. *International Journal of Telerehabilitation*, 3(1), 31-42. https://doi.org/10.5195/ijt.2011.6064

Hamilton, E. R., Rosenberg, J. M., & Akcaoglu, M. (2016). The substitution augmentation modification redefinition (SAMR) model: A critical review and suggestions for its use. *TechTrends*, 60(5), 433-441. https://doi.org/10.1007/s11528-016-0091-y

Hao, Y., Mendelsohn, J., Sterneck, R., Martinez, R., & Frank, R. (2021). Probabilistic predictions of people perusing: Evaluating metrics of language model performance for psycholinguistic modelling. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* (pp. 75-86). https://doi.org/10.18653/v1/2020.cmcl-1.10

Harris, J., Mishra, P., & Koehler, M. (2009). Teachers' technological pedagogical content knowledge and learning activity types: Curriculum-based technology integration reframed. *Journal of Research on Technology in Education*, 41(4), 393-416. https://doi.org/10.1080/15391523.2009.10782536

Herring, M. C., Koehler, M. J., & Mishra, P. (Eds.). (2016). *Handbook of technological pedagogical content knowledge (TPACK) for educators* (2nd ed.). Routledge.

Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. *Educational Technology Research and Development*, 48(3), 23-48. https://doi.org/10.1007/BF02319856

Hilari, K., & Byng, S. (2009). Health-related quality of life in people with severe aphasia. *International Journal of Language & Communication Disorders*, *44*(2), 193–205. https://doi.org/10.1080/13682820802008820

Hill, A. J., & Theodoros, D. (2017). Research into telehealth applications in speech-language pathology. *Journal of Telemedicine and Telecare*, 23(1), 37-47. https://doi.org/10.1258/135763302320272158

Hill, A. J., Theodoros, D., Russell, T., & Ward, E. (2020). Using telerehabilitation to deliver evidence-based speech-language pathology services. In R. Swisher & L. T. Goldberg (Eds.), *Innovations in allied health fieldwork education* (pp. 145-162). Springer.

Hillis, A. E. (2007). Aphasia: Progress in the last quarter of a century. *Neurology*, 69(2), 200-213. https://doi.org/10.1212/01.wnl.0000265600.69385.6f

Hilton, J. T. (2016). A case study of the application of SAMR and TPACK for reflection on technology integration into two social studies classrooms. *Social Studies*, 107(2), 68-73. https://doi.org/10.1080/00377996.2015.1124376

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Holton, D., Mackridge, P., & Philippaki-Warburton, I. (2004). *Greek: A comprehensive grammar of the modern language*. Routledge.

Honnibal, M., & Montani, I. (2017). *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*. https://spacy.io

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer*, 18(8), 500-510. https://doi.org/10.1038/s41568-018-0016-5

Hu, J., Gauthier, J., Qian, P., Wilcox, E., & Levy, R. (2020). A systematic assessment of syntactic generalization in neural language models. In *Proceedings of ACL 2020* (pp. 1725-1744). https://doi.org/10.18653/v1/2020.acl-main.158

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning. *English Language Research Journal*, 4, 1-16.

Johns, T. (1998). Contexts: The background, development and trialing of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100-115). Longman.

Johnson, D. W., & Johnson, R. T. (1999). Making cooperative learning work. *Theory Into Practice*, 38(2), 67-73. https://doi.org/10.1080/00405849909543834

Joshi, P., Santy, S., Budhiraja, A., et al. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of ACL 2020* (pp. 6282-6293). https://doi.org/10.18653/v1/2020.acl-main.560

Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed. draft). Retrieved from https://web.stanford.edu/~jurafsky/slp3/

Kambanaros, M., & van Steenbrugge, W. (2006). Noun and verb processing in Greek progressive aphasia. *Brain and Language*, 97(2), 157-167. https://doi.org/10.1016/j.bandl.2005.10.001

Karasimos, A., & Makri, V. (2026, to be publ.). AI in English morphological processing: a GPT attempt or “misapproach”? *Selected papers of International Symposium on Theoretical and Applied Linguistics 26*. Aristotle University of Thessaloniki.

Karasimos, A., Makri, V., & Petropoulou, E. (2025). AI in German and Greek morphological nominal processing: an LLMs evaluation. Presented in *International Conference on Greek Linguistics*. 23-26 September 2025. Cambridge, UK: University of Cambridge.

Kasneci, E., Sessler, K., Küchemann, S., Bonnet, M., …, Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Keck, C. S., & Doarn, C. R. (2014). Telehealth technology applications in speech-language pathology. *Telemedicine and e-Health*, 20(7), 653-659. https://doi.org/10.1089/tmj.2013.0295

Khan, M. F. (2025). Syntactic Transformation in Large Language Models (LLMs): A Transformational Generative Grammar (TGG) Perspective. *Dibon Journal of Languages*, 1(1), 24–43. https://doi.org/10.64169/djl.23

Kihoza, P., Zlotnikova, I., Bada, J., & Kalegele, K. (2016). Classroom ICT integration in Tanzania: Opportunities and challenges from the perspectives of TPACK and SAMR models. *International Journal of Education and Development using ICT*, 12(1), 107-128.

Kintz, S., & Wright, H. H. (2018). Discourse measurement in aphasia research: Have we reached the tipping point? A review and analysis of discourse measures in aphasia research from 2011–2015. *Aphasiology*, 32, 13-35. https://doi.org/10.1080/02687038.2017.1398803

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education*, 9(1), 60-70.

Koleck, T. A., Dreisbach, C., Bourne, P. E., & Bakken, S. (2019). Natural language processing of symptoms documented in free-text narratives of electronic health records: A systematic review. *Journal of the American Medical Informatics Association*, 26(4), 364-379. https://doi.org/10.1093/jamia/ocy173

Kolk, H. (1995). A time-based approach to agrammatic production. *Brain and Language*, 50(3), 282-303. https://doi.org/10.1006/brln.1995.1049

Koutsikakis, J., Chalkidis, I., Malakasiotis, P., & Androutsopoulos, I. (2020). GREEK-BERT: The Greeks visiting Sesame Street. In *Proceedings of LREC 2020* (pp. 3149-3157). https://doi.org/10.1145/3411408.3411440

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42-78. https://doi.org/10.3102/00346543155814

Kulmizev, A., Bizzoni, Y., Nivre, J., et al. (2024). Do multilingual language models capture differing moral norms? *Computational Linguistics*, 50(1), 1-34. https://doi.org/10.48550/arXiv.2203.09904

Kuzman, T., Mozetič, I., & Ljubešić, N. (2023). *ChatGPT: Beginning of an end of manual linguistic data annotation? Use case of automatic genre identification*. arXiv preprint arXiv:2303.03953. https://doi.org/10.48550/arXiv.2303.03953

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Liddy, E. D. (2001). Natural language processing. In *Encyclopedia of library and information science* (2nd ed.). Marcel Dekker.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Havashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1-35. https://doi.org/10.1145/3560815

Lombardi, M. M. (2007). Authentic learning for the 21st century: An overview. *Educause Learning Initiative*, 1(2007), 1-12.

Mita, M., Sakaguchi, K., Hagiwara, K., Mizumoto, T., Suzuki, J., & Kentaro, I. (2024). Towards Automated Document Revision: Grammatical Error Correction, Fluency Edits, and Beyond. In *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications* (BEA 2024) (pp. 251–265), Mexico City, Mexico. Association for Computational Linguistics.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.

MacWhinney, B., & Wagner, J. (2010). Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository. *Gesprachsforschung: Online-Zeitschrift zur verbalen Interaktion*, 11, 154-173.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, 25(11), 1286-1307. https://doi.org/10.1080/02687038.2011.589893

Manning, C. D. (2022). Human language understanding & reasoning. *Daedalus*, 151(2), 127-138. https://doi.org/10.1162/DAED_a_01905

Marini, A., Andreetta, S., Del Tin, S., & Carlomagno, S. (2011). A multi-level approach to the analysis of narrative language in aphasia. *Aphasiology*, 25(11), 1372-1392. https://doi.org/10.1080/02687038.2011.584690

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Mercer, N., & Howe, C. (2012). Explaining the dialogic processes of teaching and learning: The value and potential of sociocultural theory. *Learning, Culture and Social Interaction*, 1(1), 12-21. https://doi.org/10.1016/j.lcsi.2012.03.001

Miller, J. F., & Iglesias, A. (2020). *Systematic analysis of language transcripts (SALT), research version 20*. Middleton, WI: Salt Software, LLC.

Min, S., Lyu, X., Holtzman, A., et al. (2022). Rethinking the role of demonstrations: What makes in-context learning work? In *Proceedings of EMNLP 2022* (pp. 11048-11064). https://doi.org/10.18653/v1/2022.emnlp-main.759

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017-1054. https://doi.org/10.1111/j.1467-9620.2006.0068

Nanousi, V., Masterson, J., Druks, J., & Atkinson, M. (2006). Interpretable vs. uninterpretable features: Evidence from six Greek-speaking agrammatic patients. *Journal of Neurolinguistics*, 19(3), 209-238. https://doi.org/10.1016/j.jneuroling.2005.11.003

National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academies Press.

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech, Language, and Hearing Research*, 36(2), 338-350. https://doi.org/10.1044/jshr.3602.338

Nivre, J., De Marneffe, M. C., Ginter, F., Goldberg, Y., Hajič, J., Manning, C. D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R., & Zeman, D. (2016). Universal dependencies v1: A multilingual treebank collection. In N. Calzolari, K. Choukri, H. Mazo, A. Moreno, T. Declerck, S. Goggi, M. Grobelnik, J. Odijk, S. Piperidis, B. Maegaard, & J. Mariani (Eds.), *Proceedings of the 10th International Conference on Language Resources and Evaluation*, LREC 2016 (pp. 1659-1666).

O'Donnell, M. (2008). The UAM CorpusTool: Software for corpus annotation and exploration. In *Proceedings of the XXVI Congreso de AESLA* (pp. 1433-1447).

Ouyang, L., Wu, J., Jiang, X., Almeida, D., …, & Lowe, R. (2022). Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, 35, 27730-27744. https://doi.org/10.48550/arXiv.2203.02155

Papathanasiou, I., Coppens, P., & Potagas, C. (Eds.). (2017). *Aphasia and related neurogenic communication disorders* (2nd ed.). Jones & Bartlett Learning.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences*, 13(3), 423-451. https://doi.org/10.1207/s15327809jls1303_6

Pilehvar, M. T., & Camacho-Collados, J. (2019). *Embeddings in natural language processing: Theory and advances in vector representations of meaning*. Morgan & Claypool Publishers.

Prins, R., & Bastiaanse, R. (2004). Analysing the spontaneous speech of aphasic speakers. *Aphasiology*, 18(12), 1075-1091. https://doi.org/10.1080/02687030444000534

Prokopidis, P., & Papageorgiou, H. (2017). Universal Dependencies for Greek. In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies* (pp. 102-106).

Puentedura, R. R. (2006). Transformation, technology, and education [Blog post]. Retrieved from http://hippasus.com/resources/tte/

Qin, C., Zhang, A., Zhang, Z., et al. (2023). Is ChatGPT a general-purpose natural language processing task solver? In *Proceedings of EMNLP 2023* (pp. 1339-1384). https://doi.org/10.18653/v1/2023.emnlp-main.85

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347-1358. https://doi.org/10.1056/NEJMra1814259

Reeves, T. C., Herrington, J., & Oliver, R. (2002). Authentic activities and online learning. In *Quality conversations: Proceedings of the 25th HERDSA annual conference* (pp. 562-567).

Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge University Press.

Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A. (1996). The Philadelphia naming test: Scoring and rationale. *Clinical Aphasiology*, 24, 121-133.

Rochon, E., Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (2000). Quantitative analysis of aphasic sentence production: Further development and new data. *Brain and Language*, 72(3), 193-218. https://doi.org/10.1006/brln.1999.2285

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205-225. https://doi.org/10.1017/S0267190511000055

Romrell, D., Kidder, L. C., & Wood, E. (2014). The SAMR model as a framework for evaluating mLearning. *Online Learning*, 18(2), 1-15. https://doi.org/10.24059/olj.v18i2.435

Royal College of Speech and Language Therapists. (2017). *Guidance on clinical education and supervision*. Royal College of Speech and Language Therapists.

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language*, 37(3), 440-479. https://doi.org/10.1016/0093-934x(89)90030-8

Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational Researcher*, 20(3), 2-9. https://doi.org/10.3102/0013189X02000300

Schmid, M., Brianza, E., & Petko, D. (2020). Developing a short assessment instrument for Technological Pedagogical Content Knowledge (TPACK.xs) and comparing the factor structure of an integrative and a transformative model. *Computers & Education*, 157, 103967. https://doi.org/10.1016/j.compedu.2020.103967

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., et al. (2009). Anterior temporal involvement in semantic word retrieval: VLSM evidence from aphasia. *Brain*, 132(12), 3411–3427. https://doi.org/10.1093/brain/awp284

Sclar, M., Choi, Y., Tsvetkov, Y., & Suhr, A. (2023). *Quantifying language models' sensitivity to spurious features in prompt design or: How I learned to start worrying about prompt formatting*. arXiv preprint arXiv:2310.11324. https://doi.org/10.48550/arXiv.2310.11324

Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with Python. In *Proceedings of the 9th Python in Science Conference* (pp. 56-61).

Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19, 221-248. https://doi.org/10.1146/annurev-bioeng-071516-044442

Shen, C., Cheng, L., Nguyen, X-N., You, Y., & Bing, L. (2023). Large Language Models are Not Yet Human-Level Evaluators for Abstractive Summarization. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (4215–4233). https://doi.org/10.18653/v1/2023.findings-emnlp.278

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14. https://doi.org/10.2307/1175860

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1-23.

Simmons-Mackie, N., Raymer, A., Armstrong, E., Holland, A., & Cherney, L. R. (2010). Communication partner training in aphasia: A systematic review. *Archives of Physical Medicine and Rehabilitation*, *91*(12), 1814–1837. https://doi.org/10.1016/j.apmr.2010.08.026

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category: ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation* (pp. 816-820).

Stark, B. C., Meinert, K., Urena, K., Oeding, G., Fromm, D., & MacWhinney, B. (2025). Introducing the NEURAL Research Lab Data Set for Studies of Discourse and Gesture in Aphasia and Cognitively Healthy Aging Adults. *Journal of speech, language, and hearing research* (JSLHR), 68(11), 5543–5556. https://doi.org/10.1044/2025_JSLHR-24-00732

Stephany, U. (1997). The acquisition of Greek. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition* Vol. 4 (pp. 183-333). Lawrence Erlbaum.

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, 6(1), 1-10. https://doi.org/10.37074/jalt.2023.6.1.17

Thompson, C. K., Cho, S., Hsu, C. J., Wieneke, C., Rademaker, A., Weitner, B.B., Mesulam, M-M. & Weintraub, S. (2012). Dissociations between fluency and agrammatism in primary progressive aphasia. *Aphasiology*, 26(1), 20-43. https://doi.org/10.1080/02687038.2011.584691

Tucker, J. K. (2012). Perspectives of speech-language pathologists on the use of telepractice in schools: The qualitative view. *International Journal of Telerehabilitation*, 4(2), 47-60. https://doi.org/10.5195/ijt.2012.6102

Üstün, A., Bisazza, A., Bouma, G., KO, W-Y.,... & Hooker, S. (2024). *Aya model: An instruction finetuned open-access multilingual language model*. arXiv preprint arXiv:2402.07827. https://doi.org/10.48550/arXiv.2402.07827

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher-student interaction: A decade of research. *Educational Psychology Review*, 22(3), 271-296.

Varlokosta S., Stamouli S., Karasimos A., Markopoulos G., Kakavoulia M., Nerantzini M., Pantoula A., V. Fyndanis, Economou, A. & A. Protopapas (2016). A Greek Corpus of Aphasic Discourse: Collection, Transcription and Annotation Specifications. In *LREC 2016 Proceedings Workshop RAPID2016*, pp. 14-21. ISSN 2522-2686.

Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, 30, 5998-6008.

Voogt, J., Fisser, P., Pareja Roblin, N., Tondeur, J., & van Braak, J. (2013). Technological pedagogical content knowledge—A review of the literature. *Journal of Computer Assisted Learning*, 29(2), 109-121. https://doi.org/10.1111/j.1365-2729.2012.00487.x

Vyatkina, N. (2020). Corpora as open educational resources for language teaching. *Foreign Language Annals*, 53(2), 359-370. https://doi.org/10.1111/flan.12464

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, Y., Wang, L., Rastegar-Mojarad, M., Moon, S., Shen, F., Afzal, N., Liu, S., Zeng, Y., Mehrabi, S., Sohn, S., & Liu, H. (2018). Clinical information extraction applications: A literature review. Journal of biomedical informatics, 77, 34–49. https://doi.org/10.1016/j.jbi.2017.11.011

Warstadt, A., & Bowman, S. R. (2022). What artificial neural networks can tell us about human language acquisition. In J. Sprouse (Ed.), *Oxford handbook of experimental syntax* (pp. 1-44). Oxford University Press. https://doi.org/10.1201/9781003205388-2

Webster, J., Franklin, S., & Howard, D. (2007). An analysis of thematic and phrasal structure in people with aphasia: What more can we learn from the story of Cinderella? *Journal of Neurolinguistics*, 20(5), 363-394. https://doi.org/10.1016/j.jneuroling.2007.02.002

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. Journal of Science Education and Technology, 25(1), 127–147. https://doi.org/10.1007/s10956-015-9581-5

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Einashar, A., Spenser-Smith, J., & Schmidt, D. (2023). *A prompt pattern catalog to enhance prompt engineering with ChatGPT*. arXiv preprint arXiv:2302.11382. https://doi.org/10.48550/arXiv.2302.11382

M. Wilson, S., Eriksson, D. K., Brandt, T. H., Schneck, S. M., Lucanie, J. M., Burchfield, A. S., Charney, S., Quillen, I. A., de Riesthal, M., Kirshner, H. S., Beeson, P. M., Ritter, L., & Kidwell, C. S. (2019). Patterns of recovery from aphasia (Version 1). *ASHA journals*. https://doi.org/10.23641/asha.7811876.v1

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35. https://doi.org/10.1145/1118178.111821

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Woolf, B. P., Lane, H. C., Chaudhri, V. K., & Kolodner, J. L. (2013). AI grand challenges for education. *AI Magazine*, 34(4), 66-84. https://doi.org/10.1609/aimag.v34i4.2490

World Health Organization. (2001). *International classification of functioning, disability and health (ICF)*. World Health Organization.

Wu, Z., Qiu, L., Ross, A., Akyürek, E., Chen, B., Wang, B., Kim, N., Andreas, J. & Kim, Y. (2024). Reasoning or Reciting? Exploring the Capabilities and Limitations of Language Models Through Counterfactual Tasks. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1819–1862), Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.102

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 16(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., … J.-R. Wen (2025). *A survey of large language models. arXiv preprint.* arXiv:2303.18223. https://doi.org/10.48550/arXiv.2303.18223

Appendix A: Educational Materials

Automated and Manual Linguistic Annotation: LLM-Powered Corpus Analysis with CACLA	
Athanasios Karasimos	
Level of English Proficiency	This activity is for students at C1-C2 CEFR level of English proficiency (advanced undergraduate or graduate students in linguistics, computational linguistics, or NLP programs)
Course Area	This activity is designed for courses in *Introduction to Computational Linguistics*,*Natural Language Processing and Computational Techniques*, and *Corpus Linguistics and Language Teaching*.
Target Skill(s)	Corpus linguistics methodology, computational annotation techniques, morphosyntactic analysis, prompt engineering for NLP tasks, critical evaluation of LLM capabilities, collaborative computational research
Time Frame	The estimated timeframe for this activity is approximately two 150-minute lectures (5 hours total instruction time plus homework assignments)
Apps/Software needed	- LLM Software (Claude, GPT, Gemini, DeepSeek, Grok) - Web browser (Chrome, Firefox, Safari) - ELAN software (for comparison with traditional annotation tools) - Google Colab or Jupyter Notebook (optional, for advanced students) - Text editor with UTF-8 support (e.g., NotePad++, Sublime Text) - Spreadsheet software (Microsoft Excel, Google Sheets) for data compilation and statistical analysis - (Optional) Python environment with NLTK, spaCy, or similar NLP libraries for comparative analysis
Materials	- Selected transcripts from CACLA Corpora in plain text format - Linguistic annotation scheme reference guide (morphological, syntactic, lexical levels - parts of annotation schema template) - LLM prompt templates for computational annotation tasks - Gold standard annotated corpus samples for validation - Comparative analysis worksheet for LLM vs. rule-based/statistical methods - Assessment rubric for computational linguistics report - Research articles on corpus annotation methods and LLM evaluation - Greek linguistic reference materials (morphology, syntax overview, aphasia)
Overview	This lesson plan introduces Computational linguistics students to modern LLM-based approaches for linguistic corpus annotation and analysis. Using the CACLA corpora of aphasic speech as a case study, students will explore how Large Language Models can be leveraged for automated morphological, lexical, and syntactic annotation tasks. Through systematic prompt engineering, validation against gold standards, and comparative analysis with traditional computational methods, students develop critical skills in evaluating AI-driven NLP tools while understanding both their

	transformative potential and inherent limitations. The activity emphasizes the computational linguistics perspective: accuracy metrics, annotation consistency, linguistic feature extraction, and the methodological rigor required when deploying LLMs for serious linguistic research.
Aims of the Activities	At the end of this lesson plan/lecture, students will be able to: - Design and implement LLM-based pipelines for multi-level linguistic annotation- Apply prompt engineering principles to optimize linguistic analysis tasks - Extract quantitative linguistic measures from annotated corpora (TTR, MLU, morphological complexity, syntactic diversity) - Validate automated annotations systematically using gold standard comparisons - Calculate and interpret standard evaluation metrics (precision, recall, F1-score, inter-annotator agreement) - Compare LLM performance with traditional rule-based and statistical NLP methods - Identify linguistic phenomena that challenge current LLM architectures-Document computational workflows with reproducible methodologies - Critically evaluate the suitability of LLMs for different annotation tasks - Understand the special challenges of analyzing atypical (aphasic) language data - Collaborate effectively on computational linguistics research projects
Prerequisites	- Basic knowledge of linguistic analysis (morphology, syntax, semantics) - Completion of introductory corpus linguistics course or equivalent - Familiarity with Greek language structure - Basic understanding of NLP concepts (tokenization, POS tagging, parsing) - Familiarity with statistical concepts (mean, standard deviation, correlation) - Creating a free LLM account - Ability to read and understand linguistic research papers in English - (Recommended) Prior exposure to annotation tools like ELAN, UAM CorpusTool, or similar - (Optional) Experience with at least one programming language (Python preferred) at basic level
Lecture Structure (all the stages)	

Preparation	The lecturer prepared:
	- Curated selection of CACLA transcripts (8-12 samples, 200-400 words each, representing varied linguistic complexity - mainly from Greek TV Series scripts)
	- Gold standard annotations at three levels: morphological, lexical, syntactic (extracted human annotations from ELAN)
	- Comprehensive prompt template library organized by linguistic annotation task
	- PowerPoint presentation on:
	    - CACLA corpus structure and annotation scheme
	    - LLM capabilities and limitations for linguistic analysis
	    - Evaluation metrics for NLP systems
	- Video demonstration/tutorial (10-12 minutes) showing complete LLM

	annotation workflow - Detailed annotation guidelines document with linguistic examples - Comparative analysis worksheet templates - Assessment rubric for computational linguistics report - Research paper reading list on corpus annotation and LLM evaluation (Zotero, Mendeley) - Shared computational workspace (Google Drive, Microsoft One Drive) for code and data sharing - (Optional) Baseline annotation outputs from traditional NLP tools (spaCy) for comparison - Statistical analysis template spreadsheet for calculating evaluation metrics

Process	**Step 1 (Lecture 1): Introduction to Corpus Linguistics Foundations and CACLA Project (90 minutes)**
	
	- Overview presentation of corpus linguistics methodology: corpus design principles, annotation schemes, representativeness, and research applications
	- Introduce the CACLA corpora as a computational linguistics case study:
	    - Corpus composition and structure
	    - Transcription conventions and metadata
	    - Linguistic characteristics of atypical language data
	    - Annotation challenges specific to clinical corpora
	- Present the linguistic annotation task framework:
	    - Morphological annotation: lemmatization, POS tagging, morphological features (case, number, gender, tense, aspect, mood for Greek)
	    - Lexical annotation: word frequency, lexical diversity measures, lexical categories
	    - Syntactic annotation: phrase structure, dependency relations, grammaticality judgments
	- Discuss traditional computational approaches to these tasks (recap):
	    - Rule-based systems (advantages and limitations)
	    - Statistical models (CRF, HMM for sequence tagging)
	    - Neural approaches (BiLSTM, transformer-based models)
	- Divide students into groups of 3-4. Each group receives:
	    - Two CACLA transcripts (one simpler, one more complex linguistically)
	    - Digital and printed versions
	    - Annotation guidelines document
	- Groups conduct manual linguistic analysis of the simpler transcript (30 minutes- *ideally they use ELAN and an annotation template*):
	    - Identify and annotate 15-20 words with full morphological analysis
	    - Note interesting linguistic phenomena (errors, unusual structures, ambiguities) that are included in the hierarchal annotation schema
	    - Discuss annotation challenges and ambiguous cases
	- Brief group presentations (3-4 minutes each): What linguistic patterns did you observe? What annotation challenges and issues arose? How consistent were your annotations within the group?
	- Class discussion: Introduce the concept of inter-annotator agreement and why consistency matters in computational linguistics (see Varlokosta et al.,

	2016).
	
	Step 2 (Lecture 1): LLM-Based Annotation Pipeline Development (75 minutes)
	
	- Present comprehensive overview of Large Language Models for linguistic analysis:
	    - LLM architecture basics (transformer models, attention mechanisms at appropriate level)
	    - How LLMs process and generate language
	    - Documented linguistic capabilities and limitations
	    - Comparison with traditional NLP models
	- Live demonstration of LLM annotation workflow using Claude/GPT:
	    - Uploading corpus transcript
	    - Designing effective prompts for morphological annotation
	    - Requesting structured output formats (JSON, TSV, tables)
	    - Iterative prompt refinement based on output quality
	    - Extracting and processing results
	- Present principles of prompt engineering for linguistic tasks:
	    - Clarity and specificity in linguistic terminology
	    - Providing examples (from zero-shot to few-shot prompting)
	    - Specifying output format and structure
	    - Handling ambiguity and requesting confidence indicators
	    - Chaining prompts for complex multi-step analyses
	    - Prompt templates vs. dynamic prompt generation
	- Introduce the prompt template library with examples for:
	    - Morphological annotation (lemmatization, POS tagging, feature extraction)
	    - Lexical analysis (frequency, diversity, semantic fields)
	    - Syntactic annotation (constituency parsing, dependency relations)
	- Quantitative measure extraction (MLU, TTR, complexity metrics)
	- Error detection and linguistic anomaly identification
	- Groups begin computational annotation project:
	    - Each group works with both assigned transcripts
	    - Divide annotation responsibilities among group members (each takes one linguistic level)
	    - Access LLM and upload transcripts
	    - Use and adapt provided prompt templates
	    - Document all prompts used and modifications made
	    - Generate initial automated annotations
	    - Compile outputs in structured format (spreadsheet or JSON)
	- Groups share preliminary findings: *What worked? What failed? How did you refine prompts?*
	- (optional) Troubleshooting session: Addressing common issues (encoding problems, inconsistent output formats, Greek language handling)
	
	Step 3 (Lecture 2): Validation, Evaluation Metrics, and Comparative Analysis (90 minutes)
	
	- Present rigorous evaluation methodology for NLP systems:
	    - Gold standard creation and its importance

	- Evaluation metrics: precision, recall, F1-score, accuracy - Confusion matrices for error analysis - Statistical significance testing - Error categorization and analysis - Distribute gold standard annotations for the transcripts students have been analyzing - Systematic validation exercise (60 minutes): - Quantitative evaluation: *- Groups compare their LLM-generated annotations with gold standards* *- Calculate precision, recall, and F1-score for some annotation types* *- Create confusion matrices for POS tagging* *- Identify systematic error patterns* *- Use provided spreadsheet templates for metric calculation* - Qualitative error analysis: *- Categorize errors by type (linguistic level, error source)* *- Identify which linguistic phenomena the LLM handles well/poorly* *- Analyze whether errors correlate with specific Greek morphological features* *- Note whether atypical language characteristics affect accuracy* *- Examine false positives vs. false negatives* - Comparative analysis: *- Compare LLM performance across different annotation tasks* *- Compare simple vs. complex transcripts* *- Analyze time efficiency: manual vs. automated annotation* *- Consider accuracy-speed trade-offs* *- Introduce advanced measurement extraction:* *- Design prompts for calculating linguistic complexity metrics:* *- Mean Length of Utterance (MLU)* *- Type-Token Ratio (TTR) and variants (MTLD, MATTR)* *- Morphological complexity indices* *- Syntactic complexity measures (subordination index, dependency distance)* - Groups extract these measures from their annotated corpora - Compile results in shared spreadsheet for cross-group comparison - Groups complete comprehensive comparative analysis worksheets: - Strengths of LLM approach for each annotation type - Specific limitations discovered with examples - Computational efficiency analysis - Recommendations for optimal use cases - Suggestions for hybrid approaches (LLM + rule-based/statistical methods) **Step 4 (Lecture 2): Advanced Applications, Reproducibility, and Research Presentations (75 minutes)**

	- Brief teacher presentation on advanced computational linguistics applications: - Batch processing multiple corpus files efficiently - Using Claude's API for programmatic access (Python examples) - Building custom annotation artifacts with visualization - Developing hybrid pipelines (LLM + traditional NLP tools) - Reproducibility best practices in computational linguistics research - Version control for corpus annotation projects - Optional advanced exercise for interested students: - Use LLM to generate Python code for batch annotation processing - Create data visualization scripts for linguistic feature distributions - Groups finalize comprehensive computational linguistics research reports focusing on: - Methodology section: Clear description of corpus, annotation scheme, LLM configuration, prompts used, evaluation procedures - Results section: Quantitative evaluation metrics with tables and figures, error analysis with examples, linguistic measure extraction results - Comparative analysis: LLM vs. traditional methods, accuracy-efficiency trade-offs, task-specific performance differences - Discussion: Linguistic phenomena that challenge LLMs, implications for computational linguistics research, recommendations for future work - Reproducibility: All prompts documented, workflow clearly described, data and code availability stated - Each group member presents their specialized section - Class discussion synthesizing findings across all groups: - Which annotation tasks are most suitable for LLM automation? - What linguistic features consistently challenge LLMs? - How does atypical language affect LLM performance? - What role should LLMs play in computational linguistics research? - Reflection on computational linguistics methodology: - Balancing automation with linguistic rigor - Importance of validation and transparency - Future directions for LLM applications in corpus linguistics - Ethical considerations in computational analysis of clinical data
Variations	For different computational backgrounds: - *Advanced Computer Science students*: Emphasize algorithmic aspects, include API usage, develop Python scripts for batch processing, implement statistical significance testing, compare with transformer models like BERT/GPT - *Linguistics students*: Focus on linguistic analysis quality, reduce programming components, emphasize qualitative error analysis, connect to theoretical linguistics For different linguistic focus: - *Morphology-focus*: Deep dive into Greek morphological complexity, morpheme segmentation, allomorphy, paradigm analysis - *Syntax-focus*: Focus on parsing accuracy, dependency relations, constituency structures, grammaticality judgments

	- *Semantics-focus*: Lexical semantics annotation, semantic role labeling, meaning representations Alternative corpora and applications: - Replace CACLA with other corpora: child language acquisition, language disabilities data, learner corpora - Cross-linguistic comparison: Apply same methodology to parallel corpora in multiple languages Technology variations: - Compare multiple LLMs (Claude vs. GPT-5 vs. Gemini) for same tasks - Use specialized linguistic annotation platforms (WebAnno, Sketch Engline, CatMa) alongside LLMs
Troubleshooting Tips	**Technical Issues:** - *LLM access problems*: Have backup accounts ready by providing alternative LLM options - *Greek text encoding*: Ensure UTF-8 encoding throughout; test copy-paste vs. file upload; provide pre-processed clean text files - *Response length limitations*: Break long transcripts into smaller chunks; use multi-turn conversations; teach students to request continuation **LLM-Specific Challenges:** - *Hallucinated annotations*: Train students to always validate against source text and do not easily accept the output. - *Inconsistent annotation across examples*: Design prompts emphasizing consistency. - *Greek morphological errors*: Provide LLM with explicit Greek morphology reference in prompt; use few-shot examples. - *Ambiguity handling*: Request confidence scores or multiple analyses; teach students to identify genuinely ambiguous cases vs. LLM errors **Pedagogical Challenges:** - *Over-reliance on automation*: Require manual annotation baseline for comparison; grade based on critical evaluation, not just LLM output; emphasize validation importance - *Insufficient linguistic knowledge*: Provide supplementary materials on Greek morphosyntax; pair stronger/weaker students strategically. - *Time management issues*: Provide realistic time estimates for each task; prioritize core activities if running behind. - *Statistical concepts confusion*: Provide worked examples of metric calculations; offer spreadsheet templates with formulas; explain concepts with linguistic examples rather than pure math **Computational Challenges:** - *Data format issues*: Provide clear format specifications and examples; offer conversion scripts; validate formats before processing - *Evaluation metric calculation errors*: Provide spreadsheet templates with built-in formulas; show step-by-step calculations; offer troubleshooting session

Links	Primary Tools: LLMs: GPT https://chatgpt.com/ | Gemini https://gemini.google.com/app Claude https://claude.ai/ | DeepSeek https://www.deepseek.com/en | Grok https://grok.com/ CACLA Corpus Resources: https://enl.auth.gr/cacla/ Corpus Linguistics Resources: Introduction to Corpus Linguistics: https://www.corpuslinguis.org Lancaster University Corpus Linguistics Page: https://www.lancaster.ac.uk/linguistics/about/research/corpus-linguistics/ Corpus Linguistics Tutorial (YouTube): [relevant playlist] Text Encoding Initiative (TEI) Guidelines: https://tei-c.org/guidelines/ NLP and Computational Tools: ELAN Annotation Software: https://archive.mpi.nl/tla/elan Natural Language Toolkit (NLTK): https://www.nltk.org spaCy NLP Library: https://spacy.io UAM CorpusTool: http://www.corpustool.com AntConc: https://www.laurenceanthony.net/software/antconc/ Evaluation and Metrics: SciKit-Learn Metrics Documentation: https://scikit-learn.org/stable/modules/model_evaluation.html LLM and Prompt Engineering: Prompt Engineering Guide: https://www.promptingguide.ai OpenAI Prompt Engineering: https://platform.openai.com/docs/guides/prompt-engineering Anthropic Prompt Engineering Guide: https://docs.anthropic.com/en/docs/build-with-claude/prompt-engineering Few-Shot Learning with LLMs: https://www.coursera.org/learn/zero-shot--few-shot-learning-master-ai-with-minimal-data
Other Information	**Copyright and Ethics:** - All CACLA corpus materials used with appropriate authors/creators permissions and ethical approvals - LLM-assisted analysis must be clearly labeled in all reports and presentations - Students must acknowledge AI tools used: "Linguistic annotations were generated using LLM name (Company name) and validated against gold standard annotations" **Accessibility Considerations:** - Provide transcripts in multiple formats (plain text, annotated, screen-reader friendly) - Ensure all video materials have captions - Allow flexible group composition to accommodate diverse needs - Provide extended time for students who require it

	Future Development: - Activity template will be iteratively refined based on instructor feedback - Additional expansion of the CACLA corpora and linguistic phenomena may be incorporated - Evaluation rubrics may be adjusted for different course contexts

Athanasios Karasimos (akarasimos@enl.auth.gr) is an Assistant Professor in Computational Linguistics, Aristotle University of Thessaloniki, School of English. He is a graduate of the Department of Philology, University of Patras. He holds two European Masters in Speech and Language Processing (one of them at the University of Edinburgh) and his doctoral dissertation is in Computational Morphology. He participated in several research projects on Modern Greek dialects, corpora, aphasic speech, Digital Humanities, and training of English language teachers. He was a postdoctoral research fellow funded by IKY. He worked as an Adjunct Lecturer at HOU, AUTh, and NKUA teaching Educational Technology, Research Methodology, Computational Linguistics and Corpus Linguistics. He is a researcher in the national infrastructure for Digital Humanities DARIAH-GR / DYAS (Academy of Athens). His research interests focus on Computational Linguistics and machine learning, the use of corpora, education technology, and integrating video and board games into language teaching and learning.

Evangelia–Antonia Efstratiadou (e.efstratiadou@go.uop.gr) is an Assistant Professor in Speech and Language Therapy at the University of the Peloponnese. She is a graduate of the Technological Educational Institute of Western Greece with a degree in Speech and Language Therapy. She holds an MSc in Human Communication and a PhD in Language and Communication Sciences from City, University of London, where her doctoral research focused on aphasia therapy. She has over fifteen years of clinical experience in paediatric and adult neurogenic communication disorders and has worked in rehabilitation and hospital settings providing assessment, intervention, and clinical supervision. She has participated in large-scale national and European research projects, including the THALES Project, the CACLA Project, and MapCosmos. She co-leads national initiatives on aphasia assessment and implementation science, contributing to the development of LexiGrAph, a Greek lexical and grammatical aphasia assessment tool, and to the adaptation of KomTil for Greek clinical contexts. Her work has been presented at major international conferences such as IARC and ICPLA. Her research interests focus on aphasia rehabilitation, evidence-based intervention methods (such as Semantic Feature Analysis and Mapping Therapy), psycholinguistic mechanisms underlying language processing, lexical and grammatical assessment in Greek, communication partner training, implementation science in speech and language therapy, qualitative and quantitative analysis of aphasic speech, and the translation of research evidence into clinical practice. She also has extensive experience in training and mentoring healthcare professionals working with individuals with acquired communication disorders.

Dr. Christos Papatzalas (cpapatz@upatras.gr) is an Assistant Professor in Speech-Language Pathology: Communication Disorders, Department of Speech and Language Therapy, University of Patras. He holds a PhD in Clinical Speech-Language Pathology from the University of Thessaly (Faculty of Medicine, Neurosurgery Clinic), focusing on language assessment during awake brain surgery, completed with a three-year scholarship from the Greek State Scholarships Foundation (IKY). He also holds a BSc in Speech and Language Therapy and an MSc in Linguistics from the University of Patras. His research focuses on communication disorders (and bilingualism) with emphasis on aphasia, tumor-related language disorders, and intraoperative language mapping. He has participated in multiple research projects and has published in international peer-reviewed journals and conference proceedings.

Dr. Ilias Papathanassiou (ipapatha@upatras.gr) is a Professor of Speech and language Therapy, at the Department of Speech and Language Therapy, University of Patras. He has almost 30 years of experience in the area of rehabilitation voice, neurogenic communication and swallowing disorders. He was awarded his PhD from the Institute of Neurology, University College London, University of London and has done postdoctoral studies at the Department of Speech and Language Therapy at University of Queensland, Australia. He has published books (Acquired Neurogenic Communication Diosrders, 2000; The Sciences of Aphasia, 2002; Aphasia and Related Neurogenic Communication Disorders, 2012, 2016, 2022), has published dozens of scientific articles and studies and has over a hundred international participations with presentations in an international peer reviewed conferencesHe has served as a member of the editorial board of the scientific journals Aphasiology, Communication Disorders Quarterly, Clinical Archives of Communication Disorders, ASHA perspectives SIG17, and has been responsible for the book reviews of the International Journal of Language and Communication Disorders and Folia Phoniatrica et logopedica. Dr. Papathanasiou has received awards in both Great Britain and America. In 2008, the Royal College of Speech Therapists in Great Britain awarded him the honorary title of. In 2014, the American Speech Language Hearing Association (ASHA) was awarded the honorary title of Fellow (the Highest Award for Speech and Language Therapist in the USA) for its contribution to research, teaching and international organizations of the profession. He is the 2020 recipient of the ASHA Certificate of Recognition for Outstanding Contributions in International Achievement.