

ENGLISH LANGUAGE AND LINGUISTICS

Research project
Title of the presentation “LIKE + THAT-clause” structures in contemporary English: corpus design and methodological challenges
Main topic of and main issues in the PhD project (incl. the thesis proposed or hypothesis defended) My dissertation focuses on structures which follow the pattern “LIKE + THAT-clause”, as in “They like that he’s a business guy”. These relatively new structures remain a matter of debate among scholars. Some linguists argue that “LIKE + THAT-clause” is ungrammatical (Huddleston & Pullum 2002: 963), or non-canonical (Quirk et al. 1985: 1175; Sinclair et al. 1996: 95). Others, on the contrary, maintain that these structures are acceptable (Levin 1991: 191; Dixon 1991: 259). Moreover, the Corpus of Contemporary American English alone features several thousand occurrences of this structure. One of the main goals of my dissertation is therefore to assess the specificities of this structure and confirm whether it can be considered as complementation pattern of LIKE.
Methodology, theoretical perspectives and standpoints This presentation will focus on methodological considerations related to corpus design and data collection. Bearing in mind the central concern for reliability and representativeness (McEnery & Hardie 2012: 8), how can one effectively capture a structure whose core components are multifunctional? First, we propose to account for the choice of the main corpus (the COCA) and of the control corpus (enTenTen). Indeed, as a large and balanced corpus, the COCA allows for a comprehensive analysis of an emerging structure. The control corpus is used to assess the specificity of the findings and explain potential corpus effects. Then, we will present the collection strategy based on a semi-automated protocol which combines corpus queries with manual annotation and validation.
Data/corpora/materials in the PhD project COCA - main corpus enTenTen - control corpus
Results obtained so far The results obtained so far remain preliminary and will require a more comprehensive investigation based on both the main and control corpora. However, some initial findings presented at a conference already confirm that the subordinate clause satisfies the syntactic and semantic criteria of a prototypical complement clause. In addition, the emerging status of this structure has been confirmed in an article currently under review.

Research project
Title of the presentation The Language of Control: Corporate Power and Consumer Agency in South Park
Main topic of and main issues in the PhD project (incl. the thesis proposed or hypothesis defended) This paper analyses speech acts and lexis in ten South Park episodes focusing on corporate power, consumerism and predatory persuasion. It examines how corporate and institutional powers utilise language to project and abuse dominance, and how consumers and engaged citizens use speech acts to articulate powerlessness or stage resistance. The study argues that the dominant groups maintain dominance through “linguistic camouflage”, which refers to masking the link between the benefit and harm. I posit that power is exercised by weaponising language to distort reality, reframing exploitation as benevolence. Resistance is achieved through active linguistic agency – the capacity to dismantle this mask by voicing the truth versus the risk of internalising the oppressor’s narrative.
Methodology, theoretical perspectives and standpoints Speech acts are defined as utterances with distinct illocutionary force, ranging from single sentences to longer sequences. The dataset is categorised into three main types for powerful and powerless characters, specifically analysing combinations of Constatives, Directives and Commissives paired with Expressive speech acts. This classification utilises taxonomies by Bach and Harnish (1979) and Guiraud et al. (2011). Power is primarily defined by Van Dijk (2008) and Fairclough (1996), viewing language as the primary instrument of dominance. Additionally, the analysis examines lexis used to project power or resistance and groups these items into clusters based on their shared function.
Data/corpora/materials in the PhD project The analysis examines speech acts and lexis selected from the following ten episodes from the TV show South Park (dir. by Trey Parker and Matt Stone, 1997–2025): Gnomes (1998), Something Wall-Mart This Way Comes (2004), Freemium Isn’t Free (2014), Cash for Gold (2012), Black Friday (2013), Sponsored Content (2015), Truth and Advertising (2015), HumancentiPad (2011), Handicar (2014) and Japanese Toilet (2023).
Results obtained so far A stark contrast between dominance and resistance can be observed. The powerful exploit speech acts mainly to camouflage their exploitation and impose a distorted reality, for example by feigning friendship or showing benevolence. They employ a lexis of false generosity (e.g. “free coins”) and frame dissent as confusion (“feeling unsure of yourself”) to manipulate the addressee. In response, the powerless utilise speech acts to expose illusions through truth-telling, while simultaneously demonstrating internalised subjugation or lack of agency. Their lexis involves both precise naming of reality (“sponsored content”) and delusion (“valuable”, “just a game”). In conclusion, the powerful use language to obscure truth, while the powerless use it to either reclaim it or signal their defeat.

Research project
Title of the presentation Lexical availability and VLS use in English L2 and Spanish L3 in the Hungarian academic context
Main topic of and main issues in the PhD project (incl. the thesis proposed or hypothesis defended) My main topic is vocabulary acquisition in the academic context, which involves not only frequency-based academic vocabulary lists but technical vocabulary as well. Since these two types of vocabulary co-occur in academic genres, and students have to use them both in academic texts, a usage-based definition that takes into consideration all vocabulary items acquired by students in academic contexts is necessary. The main aims of my study are to explore <ol style="list-style-type: none">i. the relationship between proficiency levelii. the organization of academia-related lexical items in the mental lexiconiii. the use of vocabulary learning strategies (VLSs) in the academic context and to compare the English L2 or Spanish L3 of learners from different study years at a Hungarian university across these variables.
Methodology, theoretical perspectives and standpoints Although the relationship between VLS, lexical development and proficiency has been widely studied, it is not known yet how they are related to lexical access and the structural networks of the mental lexicon. As one of the main objectives of my study is to explore how VLS use in the academic context is related to proficiency level and data about lexical access processes yielded with a lexical availability task (LAT), the lexical fluency of academia-related cue words will be measured with variables, such as number of types and tokens produced to each centre of interest and graph metrics that provide information about the structure of the mental lexicon. Not only the correlations between all these variables will be computed, but also the cause-effect relationships with regression analyses.
Data/corpora/materials in the PhD project In the next stage, the main phase of data collection, the instruments will be an Oxford Placement Test for English and a commonly used placement test for Spanish to determine the proficiency level of the participants, a lexical availability task (LAT) containing 3 academia-related centres of interest and another general taxonomic control category, and an adapted version of a questionnaire (Chung et al., 2025) that measures VLS use in academic contexts. The fine-tuned instruments will be administered to 180 participants (30 students from years 1, 2, and 3 for both languages), and the data will be analysed in line with the aims of the study.
Results obtained so far Pilot studies have already been conducted with the second and the third instruments with the participation of about 100 students. Differences in the quantitative lexical production and in the structural organization in L1, L2 and L3 were found, and the centres of interest to use in the main study are under consideration. The adapted questionnaire was found to be a reliable instrument with high Cronbach's alpha values in most scales. However, changes will be made to improve the internal consistency of the instrument for the main study.

Silvia Dragomir (University of Bucharest, Romania)

Research project
Title of the presentation Role Nouns and Gender Stereotypes: A Study on English and Romanian
Main topic of and main issues in the PhD project (incl. the thesis proposed or hypothesis defended) My PhD project sets out to investigate the translation of role nouns from English into Romanian, aiming to shed light on the use of the masculine form when referring to women, and taking into account the role played by gender stereotypes in this process. The thesis seeks to address the following issues: 1) to what extent does English, as a lingua franca, influence a grammatical gender language such as Romanian in the use of the ‘neutral’ masculine forms for role nouns? 2) if the masculine form is used, does this happen only in the case of stereotypically masculine, prestigious professions (e.g.: president, politician), or is it also the case of stereotypically feminine, less prestigious role nouns (e.g.: nurse, secretary)?
Methodology, theoretical perspectives and standpoints In the debate concerning gender-inclusive language, feminization is an important strategy proposed by linguists to eliminate gender bias, making women more visible through language. There are many studies in Romance and Germanic languages which demonstrate that, despite intentions of neutrality, the value of the generic masculine is, most of the times, purely masculine. Experimental studies with native Romanian participants will allow us to investigate the masculine/feminine dichotomy from a synchronic perspective. In order to evaluate how/if these issues have changed over a short span of time, a corpus study will be developed, offering a diachronic perspective.
Data/corpora/materials in the PhD project For the synchronic perspective, we will use data obtained from several experimental studies (translation tasks) which will be designed in Google Forms. With respect to the diachronic study, we will design a corpus consisting of female authors’ descriptions on different book covers in the last 15 years, published in Romania. In this way, we will be able to examine the changes which took place in the translation of role nouns in light of the changes concerning linguistic norms and social perceptions over the last 15 years.
Results obtained so far Four preliminary experimental studies were conducted for my MA thesis. The results confirm that the feminine form is present in the language, but its use is restricted for reasons which are not related to structural concerns, such as gender stereotypes or the prestige of the profession. Additionally, a small corpus study was designed by my coordinator (prof. Ruxandra Vişan) and me. The corpus we analyzed consisted of 77 books translated from English to Romanian, published by Trei Publishing house between 2010 and 2025. The results were in line with previous findings, but due to a limited number of role nouns, the fluctuation between the masculine/feminine dichotomy needs to be further investigated.

Research project
Title of the presentation AI-Powered Automated Feedback Tools in EFL Writing: Learners' Perceptions and Linguistic Impact
Main topic of and main issues in the PhD project (incl. the thesis proposed or hypothesis defended) The main topic of my PhD project is the role of AI-powered automated writing tools (e.g. grammar and style checkers) in English as a foreign language (EFL) writing, with a focus on their linguistic impact on learner texts and learners' perceptions of automated feedback. The project examines how the use of MS Editor during a class intervention influences the writing process and learner engagement with corrective feedback. The project defends the hypothesis that learners perceive automated feedback more positively than teacher-provided feedback and that guided exposure to automated feedback leads to measurable improvements in writing quality.
Methodology, theoretical perspectives and standpoints Methodologically, the project is an intervention study with quasi-experimental features and adopts a mixed-methods approach. It combines quantitative and qualitative linguistic analyses of student texts from pretests and posttests with questionnaire data on learners' attitudes and self-reported practices, complemented by qualitative analysis of focus group interviews exploring learner perceptions of automated writing tools and feedback even further. The study is grounded in process-oriented approaches to second language writing and research on automated written corrective feedback, emphasising learner engagement, noticing, and reflective use of feedback within guided instructional contexts.
Data/corpora/materials in the PhD project The primary dataset consists of EFL learner written texts collected in a classroom-based pretest–posttest design. Learners produce texts without the use of automated writing tools, which allows for direct comparison of linguistic features before and after instructional intervention involving the use of an AI-powered writing tool (MS Editor). These texts form a learner corpus used for quantitative and qualitative linguistic analysis. Additional data include questionnaire responses on learners' attitudes and self-reported practices related to automated writing tools, as well as audio-recorded focus group interviews exploring learners' perceptions of automated feedback.
Results obtained so far Preliminary results from the pilot study indicate that learners frequently use AI-powered automated writing tools primarily for grammar correction, spelling, and stylistic support. Questionnaire data suggest generally positive learner attitudes, with many students reporting increased confidence and reduced anxiety when writing in English. Linguistic analysis of pretest and posttest texts shows modest improvements in text complexity. Focus group data further suggest that while learners value automated feedback, many engage with it at a surface level, highlighting the need for pedagogical guidance to support deeper linguistic processing.

Research project
Title of the presentation Who Benefits from Learning Through English? Gender, SES and Extramural Exposure in CLIL
Main topic of and main issues in the PhD project (incl. the thesis proposed or hypothesis defended) My PhD project investigates English learning gains among young learners (YLS; aged 10–12) in Content and Language Integrated Learning (CLIL) programmes, where a subject such as Science is taught through English. The study examines how individual and contextual factors influence learners' linguistic development. Drawing on previous research (e.g. Soto-Corominas et al., 2024; Pérez Cañado, 2020), the project focuses on three key variables: (1) gender, (2) socioeconomic status (SES), and (3) exposure to extramural English (EE). It aims to explore how these factors shape learners' English proficiency and contribute to differential learning outcomes in CLIL contexts.
Methodology, theoretical perspectives and standpoints The study adopts a quantitative design. Based on the amount of CLIL exposure, learners are assigned to three groups: non-, low-, and high-CLIL. Participants meet predefined inclusion criteria and complete questionnaires on language background, EE exposure, and SES. English proficiency is assessed in listening, reading, speaking, and writing using standardised and validated tests. - Gender effects in CLIL are assumed to be levelled out (e.g. Heras & Lasagabaster, 2015). - Findings on SES remain mixed, with CLIL shown to both mitigate (Lorenzo et al., 2021) and reproduce inequalities (Dallinger et al., 2016). - EE is expected to support language development depending on type and intensity of exposure (Sundqvist, 2009).
Data/corpora/materials in the PhD project The study draws on several data sources. These include questionnaires collecting information on learners' language background, SES, and exposure to EE. The SES questionnaire gathers data on parents' educational attainment, occupational status and type of employment, family possessions, and cultural capital. EE exposure is measured through a self-report questionnaire covering different activity types (e.g. watching, reading, listening, online surfing), with learners indicating the amount of time spent on each activity in hours per week. These data are used to contextualise learners' linguistic development and to examine the relationship between background variables and English learning outcomes in CLIL and non-CLIL settings.
Results obtained so far Gender: Findings suggest that the gender gap reported to narrow in CLIL contexts could not be fully confirmed. Girls tend to outperform boys in reading and writing in high-intensity (HI) CLIL settings. This indicates that HI compared to low-intensity (LI) CLIL may be especially beneficial for females, while boys show gains only in reading. SES: Results partially support a levelling effect. Girls with lower-SES outperform higher-SES peers when enrolled in HI strands, suggesting that intensive CLIL may mitigate SES disparities. EE: Learners in HI CLIL strands tend to achieve higher pronunciation scores than those in LI contexts. However, the amount of EE alone shows only a weak association with pronunciation, indicating that qualitative aspects of exposure may play a more decisive role.

Dalia Mennella (University of Naples “L’Orientale”, Italy)

Research project
Title of the presentation The Chinese Stereotype in US Cinema: A Multimodal Discourse Analysis from the 1910s
Main topic of and main issues in the PhD project (incl. the thesis proposed or hypothesis defended) The Chinese stereotype in American cinema did not emerge out of nowhere, its roots lie in historical and social motivations. Throughout the past century, the Chinese people has been repeatedly depicted as an obnoxious, non-conventional, transnational “other” through the lens of American narrations, which are filled with stereotypical images of a China that swings between being an idyllic Promised Land or a hellish place. To date, critical scholarship defines the history of the Chinese stereotype in American cinema with the image of an oscillating pendulum (Greene, 2011) within a binarism that “feeds and reflects the fundamental divide between East and West between the (American) self and the (Chinese) other (Greene: 2014, 11).
Methodology, theoretical perspectives and standpoints Building on the theoretical frameworks proposed by Wong (1978) and Rey Chow (1998) which foreground that Eurocentric representations of China in cinema have failed in accurately representing the heterogeneity of each and every Asian culture, this research proposal concentrates on a corpus of US films spanning from the 1910s to the contemporary era. The analysis employs a combined methodological approach integrating Multimodal Discourse Analysis, Functionalist Stylistics and Pragmatics, to explore both new and old stereotypes. The analysis adopts both a quantitative and a qualitative approach as these will allow for a truly comprehensive analysis based on a larger body of relevant data, aimed at drawing out recurring representational patterns and their diachronic evolution in US cinema.
Data/corpora/materials in the PhD project This research proposal concentrates on a corpus of US films from the 1910s to the contemporary era. Each of the selected cinematic resources belongs to different genres, yet their common denominator lies in the portrayal of the Chinese nation to the Western audience. The selection of the corpus has been made through the comparison of several compendiums in which some resources have been ignored or just superficially mentioned in favour of a more thorough analysis of others. Others, what is more, will be explored at this specific occasion, as at the current time of writing, they have not been released yet.
Results obtained so far An analysis grounded in Halliday’s transitivity system and Multimodal Discourse Analysis produced preliminary findings highlighting differences in the construction of otherness between older and recent films. While the former overtly reproduce stereotyping narratives through verbal language, the latter do so in a subtler yet pervasive form. Earlier texts depict China as underdeveloped. Contemporary productions portray Chinese people as victims of the CCP, shifting stereotyping from ethnicity to political allegiance. Modern China is framed as a threat, supplying the narration of a modern yellow peril (Marchetti, 1993). Visual and sound strategies reinforce a racialised imaginary through orientalist tropes (Said, 1978), contributing to an enduring project of othering.

Research project
<p>Title of the presentation</p> <p>Semiotics and its Application in Fashion Discourse Aesthetic</p>
<p>Main topic of and main issues in the PhD project (incl. the thesis proposed or hypothesis defended)</p> <p>The first part of my PhD thesis, that makes the subject of my presentation, investigates how semiotics and figurative language such as metaphor, personification, irony, and hyperbole, shape the aesthetic, cultural, and rhetorical dimensions of fashion discourse, particularly in online fashion criticism. Addressing a significant research gap, I propose a new tool Clothing Aesthetic Description (CAD) framework to systematically analyze how visual elements are translated into persuasive and culturally meaningful language. The thesis contends that figurative language is central to constructing identity, cultural myths, and influence consumer perception in contemporary fashion media, making fashion discourse a powerful site of meaning-making and social influence.</p>
<p>Methodology, theoretical perspectives and standpoints</p> <p>Methodology: Qualitative discourse analysis of 43 online fashion reviews and articles, using thematic and rhetorical analysis. Figurative language is identified and coded with an adapted Metaphor Identification Procedure. The original Clothing Aesthetic Description (CAD) framework maps visual fashion elements into language. Theoretical perspectives: Semiotics (Barthes), intersemiotic translation (Jakobson), conceptual metaphor theory (Lakoff & Johnson), and critical discourse analysis (Charteris-Black), with insights from aesthetic philosophy. Standpoints: Interdisciplinary, culturally focused, and attentive to the shift from traditional to digital media in fashion discourse.</p>
<p>Data/corpora/materials in the PhD project</p> <p>A corpus of 43 online fashion reviews: This includes both video transcripts and written articles from platforms such as YouTube (e.g., HauteLeMode /Luke Meagher, Sabina Teodora), The Guardian, and Elle Romania. Red-carpet event commentary: Analyses focus on major events like the Oscars, Fashion Weeks, and other high-profile occasions. Fashion blogs and online resources: Used to define and contextualize specific fashion aesthetics (e.g., cottagecore, regalcore, preppy chic). Audience comments and engagement metrics: To assess public reception and the persuasive effect of figurative language. Annotated excerpts: Selected for their use of metaphors, irony, and other stylistic devices, coded using the Clothing Aesthetic Description (CAD) framework.</p>
<p>Results obtained so far</p> <p>Figurative language is highly prevalent in online fashion criticism, with metaphors, personification, irony, and hyperbole commonly used to frame clothing aesthetics. The Clothing Aesthetic Description (CAD) framework effectively identifies and categorizes both the denotative (physical) and connotative (symbolic, cultural) elements of fashion discourse. Online fashion platforms exhibit diverse rhetorical and semiotic strategies, reflecting differences in cultural context, platform, and target audience. The language used by critics and influencers strongly influences consumer interpretation, desirability, and the perceived value of fashion items. The analysis confirms that digital fashion discourse operates as a powerful site of identity construction and cultural meaning-making.</p>

Research project
Title of the presentation The Construction of Childlessness through Discourse: a Corpus-Based Approach
Main topic of and main issues in the PhD project (incl. the thesis proposed or hypothesis defended) The main focus of the study is to investigate the dominant issues and ideological assumptions that shape representations of childlessness in written journalistic texts, with particular attention to the ways in which gender, normativity, and social expectations are reproduced through language. The project advances the hypothesis that childlessness is frequently framed through moralising discourses, which contribute to the marginalisation of non-parenthood and the reinforcement of pronatalist social norms.
Methodology, theoretical perspectives and standpoints My project follows a corpus-based design, which enables the study of both circulating societal narratives and more formalised institutional framings, allowing for comparison across genres and communicative settings. Such comparison will be based both on quantitative (Corpus Linguistics) and qualitative (Faircloughian Critical Discourse Analysis) methods. One of the main techniques that will be used is keyword analysis.
Data/corpora/materials in the PhD project The data under consideration consist of texts obtained from the NOW (News on the Web) corpus, taken from 2010 until mid-2025, in which childlessness is explicitly discussed, including media coverage, online public discourse, and relevant institutional or policy-facing materials. The NOW corpus will be compared to the BNC2014 (British National Corpus) as a reference corpus to ensure more reliable results. The reference corpus is not expected to contain topic-relevant discourse; its role is contrastive normalisation.
Results obtained so far At this stage of the project, initial analyses have identified recurrent discursive tendencies, including the frequent problematisation of childlessness, gendered asymmetries in attribution and responsibility, and the persistence of pronatalist assumptions embedded in evaluative language. These preliminary findings support the broader argument that childlessness is not merely described but actively constituted as a socially meaningful category through discourse.

Research project
Title of the presentation Comparing NMT and LLMs in English-Italian machine translationese of news articles
Main topic of and main issues in the PhD project (incl. the thesis proposed or hypothesis defended) This project examines machine translationese (MTese) (Daems et al. 2017) in news translation from English into Italian. While MTese has been widely studied in neural machine translation (NMT), the linguistic features produced by large language models (LLMs) have only recently garnered scholarly attention (Kong & Macken, 2025), and the English-Italian pair remains largely unexplored. The project compares MTese produced by NMT and LLMs across news genres and examines how raw machine-translated output differs from comparable original Italian news, investigating potential tendencies towards standardisation. The project therefore provides empirical data on the features of MTese with implications for translator training, post-editing, computational modelling, and AI literacy.
Methodology, theoretical perspectives and standpoints The project is grounded in Corpus-based descriptive Translation Studies. Rather than aligning ST and TT, an often unfeasible approach in news translation (Davies 2014), each corpus is analysed independently to identify specific traits of MTese produced by an NMT and an LLM system and compare them with authentic Italian news. Building on earlier MTese studies (De Clercq et al. 2021), the analysis combines quantitative measures with a qualitative analysis of a small sample. By focusing on the final product rather than the translation process, the project highlights the linguistic features of texts as they circulate in real communicative contexts, reflecting the growing prominence of MT output that users encounter directly, often with minimal post-editing (Trope & Marchan 2017).
Data/corpora/materials in the PhD project The empirical analysis is based on four corpora, combining existing corpora with MT translations. These include: (i) an English ST corpus of published news texts deriving from two existing corpora, the Dutch Parallel Corpus (Macken et al. 2010) and the EMA corpus (Molino 2010), containing a balanced number of words for news reports and opinion articles ($\approx 340,000$ words); (ii) an Italian TT corpus translated with NMT; (iii) an Italian corpus translated with LLMs; and (iv) a comparable corpus of non-translated Italian news texts (PUNT-TT corpus, $\approx 220,000$ words to be expanded). This design enables systematic comparison across genres (news vs. opinion) and systems (NMT vs. LLMs), treating machine-translated texts as independent linguistic products comparable to original Italian news discourse.
Results obtained so far At present, the project is at an early stage and has focused on literature review and research design. A scoping review shows that NMT still struggles with several features of news texts, including cultural references, idiomatic and colloquial language, long and complex sentences, cohesion, and semantic nuance. It also reveals that both the English-Italian language pair and the application of GenAI to news translation remain unexplored. A preliminary experimental study on ChatGPT translations of news texts about hate speech showed comparable lexical richness and good morphological adequacy in the TL, but also longer sentences than in the ST and slight discursive shifts. These findings informed the decision to include lexical diversity and sentence length in the project's empirical analysis.

Research project
Title of the presentation Emphasis, Speech Acts and Character Identity
Main topic of and main issues in the PhD project (incl. the thesis proposed or hypothesis defended) My PhD investigates how syntactic emphasis in English literary dialogue (e.g. it-clefts, emphatic do, tag questions) contributes to illocutionary force and the construction of character identity, and how these effects are altered in Romanian translation. I hypothesise that because English and Romanian encode structural emphasis differently, translations often preserve propositional meaning but modify (or even lose altogether) speech-act force, which in turn affects how characters' relations, attitudes and identities are perceived.
Methodology, theoretical perspectives and standpoints The study combines corpus linguistics, pragmatics and translation studies. English literary dialogue is extracted from corpora using CQL queries in Sketch Engine to identify emphatic instances (such as it-clefts, emphatic do and question tags). These are analysed pragmatically as carriers of focus and illocutionary force. Each instance is compared with its Romanian translation to examine shifts in speech-act type, politeness and facework. These shifts are then interpreted in terms of character identity and interpersonal relations.
Data/corpora/materials in the PhD project The data consist of parallel English-Romanian literary texts containing extensive dialogue, primarily from modern and post-modern fiction, where spoken-like interaction is prominent. English texts are processed in Sketch Engine, and their Romanian translations are analysed using both close and distant reading. The focus is on dialogue, as this is where emphasis, speech acts and identity construction are most visible.
Results obtained so far I have developed taxonomies of emphatic structures in English and Romanian and implemented CQL queries in Sketch Engine that successfully retrieve it-clefts, emphatic do and tag questions (and other emphatic structures) from literary dialogue. These methods were tested and presented at the DoCo conference, showing that corpus tools can detect syntactic emphasis. I have also established a theoretical link between emphasis, focus and illocutionary force, forming the basis for analysing identity and relational meaning in translation.