



VILNIUS UNIVERSITY
FACULTY OF PHILOLOGY

Khaled Ahmed

Programme of English Studies (Linguistics)

Revisiting Expressions of Stance in British English: BNC1994 vs BNC2014

Master thesis

Academic Supervisor: Assoc. Prof. Dr Rita Juknevičienė

VILNIUS
[2026]

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Assoc. Prof. Dr Rita Juknevičienė, whose patience, diligence, and scholarly guidance were instrumental throughout the writing of this thesis. Her insightful feedback and consistent support were invaluable in bringing this project to completion.

I am also deeply grateful to the Faculty of Philology at Vilnius University, particularly the Head of the English Department, as well as to the coordinators and administrators, whose dedication to supporting students throughout their academic journeys is greatly appreciated.

I wish to extend my thanks to my BA professors, Drs. Ahmed Helal, Ahmed Siddiq, and Abeer El Attar, whose instruction cultivated my interest in linguistics and laid the foundation by igniting the spark of linguistics research during my studies.

Special acknowledgement is due to Professor Dr Noha Faisal and to Amir Atef, whose assistance, encouragement, and patience throughout this process were greatly valued.

Finally, I am profoundly grateful to my family, without whose unwavering support and love this work would not have been possible.

Abstract

This MA thesis represents a corpus-based study comparing expressions of stance in two versions of the British National Corpus: BNC1994 and BNC2014. It focuses particularly on the classification of attitudinal and epistemic stance adverbs, following Douglas Biber's theoretical framework (Biber and Finegan, 1988; Biber et al., 1999; Biber, 2006). The two subcorpora under investigation are academic prose and e-language registers of the BNC1994 and BNC2014, with the aim of qualitatively analysing stance adverbials explicitly mentioned in the framework, according to specific selection criteria explained in the study, extracting 500 randomised concordance lines for each marker, totalling 2500 lines per corpus. In total, 8,242 concordance lines were manually examined across corpora to identify epistemic and attitudinal stance adverbials, using frequency rates and semantic analyses. Chi-square and Cramér's V tests were used to measure the significance of differences in the frequencies of the adverbials in the two versions of the BNC. P-values suggest significant results across corpora. The results suggest a decline in attitudinal stance expressions in academic prose registers, while epistemic markers remained almost stable. E-language registers exhibited a higher frequency of stance markers across all types. Cross-register comparisons demonstrate register-specific shifts across corpora. Grounded in the classification of stance adverbials by Biber et al. (1999), the novelty of the study lies in the manual analysis of each concordance line to determine its classification as epistemic or attitudinal semantically. These findings are interpreted within the frameworks of Halliday's (2014) register theory and Traugott's (1989) subjectification theory. The study provides insights for linguistics scholars, ESL teachers, and students.

Anotacija

Šis magistro darbas pristato korpusu grįstą tyrimą, kuriame lyginamos pozicijos raiškos dvejose British National Corpus versijose: BNC1994 ir BNC2014. Tyrime daugiausia dėmesio skiriama vertinamųjų ir episteminių pozicijosrieveiksmių klasifikacijai, remiantis Douglas Biber teoriniu modeliu (Biber ir Finegan, 1988; Biber et al., 1999; Biber, 2006). Tiriama du BNC1994 ir BNC2014 subkorporai: akademinės prozos ir elektroninės kalbos registrai. Tyrimo tikslas – kokybiškai analizuoti pozicijosrieveiksmius, aiškiai nurodytus teoriniame modelyje, pagal tyrime paaiškintus atrankos kriterijus. Kiekvienam žymeniui buvo atrinkta 500 atsitiktinių konkordanso eilučių, iš viso po 2500 eilučių kiekviename korpuse. Taikant dažnio rodiklius ir semantinę analizę, iš viso rankiniu būdu buvo išanalizuotos 8242 konkordanso eilutės, siekiant surasti episteminius ir vertinamuosius pozicijosrieveiksmius. Siekiant įvertinti skirtumų tarprieveiksmių dažnių abiejose BNC versijose reikšmingumą, buvo naudojami chi kvadrato ir Craméro V testai. P reikšmės rodo statistiškai reikšmingus skirtumus tarp korpusų. Rezultatai rodo vertinamosios pozicijosraiškų mažėjimą akademinės prozos registruose, o episteminiai žymenys išliko beveik stabilūs. Elektroninės kalbos registruose nustatytas didesnis pozicijos žymenų dažnis visų tipų kategorijose. Skirtingų registrų palyginimai atskleidžia registrams būdingus pokyčius tarp korpusų. Remiantis Biber et al. (1999) pozicijosrieveiksmių klasifikacija, tyrimo naujumas slypi kiekvienos konkordanso eilutės rankinėje analizėje, semantiškai nustatant jos episteminę arba vertinamąją klasifikaciją. Šie rezultatai interpretuojami remiantis M. A. K. Halliday registro teorija (2014) ir Elizabeth Traugott subjektyvizacijos teorija (1989). Tyrimas pateikia įžvalgų kalbotyros tyrėjams, ESL mokytojams ir studentams.

Tables and Figures

Tables

Table 1: frequency distribution normalised pmw of epistemic and attitudinal adverbs.....	28
--	----

Figures

Figure 1: Academic Prose: AP1994 vs AP2014 comparison (normalised pmw).....	34
Figure 2: Academic Prose: AP1994 vs AP2014 comparison in percentage (normalised pmw)...	34
Figure 3: E-language: EL1994 vs EL2014 comparison (normalised pmw).....	35
Figure 4: AP1994 vs EL1994 comparison (normalised pmw).....	36
Figure 5: AP2014 vs EL2014 comparison (normalised pmw).....	37
Figure 6: overall normalised distribution across corpora.....	38
Figure 7: heatmap of normalised stance adverbials distribution across corpora.....	39

Table of Contents

1. Introduction.....	7
1.1 Research Questions	8
2. Literature Review	
2.1 Briefly Introducing BNC.....	9
2.2 Evaluation in Language.....	9
2.3 Stance: definition, use, and examples	10
2.4 Stance across other frameworks.....	11
2.5 Previous Research.....	15
3. Data and Methods	
3.1 Theoretical framework.....	27
3.2 Corpus Description.....	28
3.3 Data Extraction and Annotation.....	28
3.4 Data Analysis Procedure.....	30
3.5 Inclusion and Exclusion Criteria.....	31
3.6 Reliability and Validity measures.....	34
4. Results and Discussion	
4.1 Academic Prose: AP1994 vs AP2014 (Normalised)	34
4.2 E-language: EL1994 vs EL2014 (Normalised).....	36
4.3 Different registers comparison across corpora	
4.3.1 AP1994 vs EL1994 (Normalised).....	37
4.3.2 AP1994 vs EL2014 (Normalised).....	38
4.3.3 Overall Normalised Distribution Across Sub-corpora.....	39
4.4 Discussion.....	40
5. Conclusion.....	43
Summary.....	45
References.....	48
Appendices	51

1. Introduction

The definition of corpus linguistics is controversial among scholars; whether it is a methodology or a separate branch of linguistics is an ongoing debate. However, it is safe to argue that “corpus linguistics is not a branch of linguistics, it is a methodology” (Sinclair, 1991:4). This methodology’s main focus is “on a set of procedures, or methods, for studying language” (Allan, 2013:727).

Corpus linguistics enabled the tracing of linguistic evolution synchronically and diachronically through technological advancements, which make treating very large sets of data manageable. That, in turn, expanded the corpus-based empirical research, making the study of the linguistic change possible and, most importantly, meaningful. Diachronic corpus-based research addresses basic questions about the mechanisms, rates, and trends of linguistic change that cannot be adequately answered through synchronic description alone.

This study represents a new study in corpus linguistics. It draws on Biber and Finegan (1988), Biber et al. (1999), and Biber’s (2006) framework of stance expressions, specifically stance adverbials. The study combines both quantitative and qualitative research methods. On the one hand, the qualitative approach involved manual analyses of the stance markers in order to classify them into their relevant categories: epistemic and attitudinal. On the other hand, the quantitative approach prioritises stance frequencies, demonstrating the trend of each stance across various registers within the relevant time period of each corpus, allowing further study of diachronicity and contrasting the two time periods by understanding the patterns and trends underlying stance expressions. Most of the previous literature do not collectively address all the perspectives addressed in this study: adverbial stance markers extracted from the two BNC versions, the diachronic dimension of stance change across the two time periods of the corpora, operationalization of two research approaches, qualitative and quantitative methods, and anchoring the study to Traugott’s (1989) model of subjectification, and Halliday and Matthiessen’s theory of register (2014).

1.1 Research Questions

This thesis aims to answer three research questions:

1. What are the differences in the use of stance adverbials across matched registers in BNC1994 and BNC2014?
2. How do the changed patterns vary by different registers?
3. How can the findings be interpreted in the light of language evolution theories?

To answer these questions, this study examines corpus-based studies on stance adverbials, register variation, and diachronic change in British English, and evaluates methods for comparing and analysing the corpora. The Literature review is framed around three interconnected axes: firstly, stance as a theoretical and analytical construct, secondly, research based on the British National Corpus (henceforth: the BNC), and lastly, diachronic investigations of linguistic change. Collectively, these works demonstrate the extent to which stance can be identified, retrieved, quantified, classified, and interpreted analytically across various discursive genres, while revealing methodological and rhetorical gaps that would clarify why this dissertation is important.

2. Literature Review

2.1 Briefly introducing BNC

Various corpora projects have been, and continue to be, conducted by many linguistic teams around the globe. Among them, the British National Corpus is one of the most reliable and well-known corpora projects worldwide. It is a crucial source of British English corpora studies conducted by the linguistics team at the University of Lancaster, United Kingdom. It represents British English in both written and spoken forms. There are two versions of the BNC. The first one is that of 1994, which was “compiled in the early 1990s...was uniquely ambitious in attempting to achieve a balanced and representative corpus of written and spoken British English” (Hunston, 2022, p. 248). The second version of BNC is the 2014 version (released in 2018, named after the year the project started). It contains millions of words of spoken and written English. This version was collected by Lancaster University and Cambridge University Press. It represents a new resource for research and teaching on contemporary British English. It is the successor to the original [British National Corpus](#) of 1994, which was compiled in the early 1990s. The British National Corpus 2014 (BNC2014) is a major project led by Lancaster University (<http://corpora.lancs.ac.uk/bnc2014/>). Being finite in size is a very important characteristic that allows scholars to adapt their research objects to the data provided by BNC, unlike the COBUILD project, which is built up as a monitor corpus: a type of corpus that keeps collecting linguistic samples without stopping. Both versions of the BNC contain 90% of their content as written data, with only 10% from spoken language. (Leech, Rayson, and Wilson, 2014:1). As per the LancsBox X (v 5.5.1) software, which makes the two versions of the BNC available on computers locally, the BNC1994 has ten registers of language, including academic prose, e-language, fiction, formal speech, informal speech, magazines, newspapers, other, other informative, and written-to-be-spoken. Whereas the BNC2014 has slightly different eight sub-corpora, including academic prose, e-language, fiction, informal speech, magazines, newspapers, official documents, and written-to-be-spoken (LancsBox X, v 5.5.1). As such, the BNC, with its two versions, can be considered a linguistic tracker that traces changes in British English, leaving them to linguists and analysts, who, in turn, use its data to spot patterns and explain linguistic trends.

2.2 Evaluation in language

Evaluation in linguistics refers to the ways the speakers and writers of a given language express their attitudes. Evaluation is an umbrella term that is used to describe “the expression of the speaker or writer’s attitude or stance towards, viewpoint on, or feelings about the entities or propositions that he or she is talking about. That attitude may relate to certainty or obligation or desirability or any of a number of other sets of values” (Hunston, 2000:5). The importance of the study of evaluation in language is based on three main factors. Firstly, it expresses the speaker’s or writer’s opinion, in that

it reflects the “value system” of the speaker and their community. Secondly, it maintains a relation between the speaker and the hearer, or between the writer and the reader. Lastly, it organises the discourse (ibid:6).

It is important to draw the reader’s attention to the fact that evaluation is not the only term used to describe an attitude or a stance. Hunston sheds light on this by elaborating that:

“It is clear that a broad range of terminology is used to refer to similar areas of language use: appraisal, stance, sentiment, evaluative, attitudinal or affective language, metadiscourse and evaluation, to name only the most obvious.” These terms have common grounds as these terms are “largely equivalent” with some differences that led to the “divergence in terminology”. (Hunston, 2011:10)

In this research, expressions of stance are the main focus of elaboration, more specifically, adverbial expressions of stance.

2.3 Stance: definition, use, and examples

Hunston (2011) highlights that evaluation and stance are two overlapping terms, but they are not the same. She elaborates that :

“Evaluation’ is an action—something which a person does. The action may be purely private and unexpressed (in which case it lies outside a study of language) or it may be performed in language.... ‘Evaluation’, or its near synonyms, is the set of words and phrases which express evaluative meaning” (ibid., 2011:10)

Conrad and Biber consider stance as a “cover term for the expression of personal feelings and assessments” (Conrad and Biber, 2000:57). Whereas stance, as defined by Biber and Finegan (1988), means “the overt expression of an author’s or speaker’s attitudes, feelings, judgments, or commitment concerning the message”. It is important to note that:

“A broad range of functions falls under the umbrella of ‘stance,’ including expression of certainty, generalization, and actuality. All of these express some aspect of speakers’ (or writers’) attitudes towards their messages, as a frame of reference for the messages, an attitude toward or judgment of their contents, or an indication of the degree of commitment towards their truthfulness” (Biber and Finegan, 1988)

Through his academic works on stance, Biber collaborated with many linguists. In his early works, Biber collaborated with Edward Finegan on a research paper in 1988. The two linguists constructed six subcategories of stance adverbials based on their semantic field. Adverbials were grouped solely based on their literal meaning. It is worth mentioning that the scholars’ categorisation was based mainly on the “subcategories of disjuncts described in Quirk et al. (1985, pp.612ff)” (ibid). The six subcategories of stance based on their semantic field as per Biber and Finegan (1988), are “*honestly* adverbials” expressing manner of speaking (e.g. *Personally, strictly*), “*generally* adverbials” expressing approximation, generalization (e.g. *Probably, roughly*), “*surely* adverbials” expressing conviction or certainty (e.g. *Unquestionably*), “*actually* adverbials” expressing actuality and emphasis (e.g. *really, actually*), “*maybe* adverbials” expressing possibility or likelihood, and “*amazingly* adverbials” expressing attitudes towards the content (e.g. *fortunately, ironically*) (ibid).

In their extensive 1999 research, Biber et al. (1999) defined adverbials as elements of clauses with three major functions. Firstly, they add to the circumstantial dimension of the proposition in the clause. Secondly, they express the speaker/writer's stance towards the clause. Lastly, they link the clause or some part of it to another part of the discourse. (Biber et al., 1999:672). Furthermore, Biber et al. (1999) divided the stance into three major categories: epistemic, attitudinal, and style. Firstly, epistemic stance expresses certainty (or doubt), limitations, or reliability of a proposition, including commenting on the source of information. Several sub-classes can be identified under the epistemic stance; for example, indicating the degree of certainty or doubt concerning the proposition, commenting on the actuality or the reality of the proposition, indicating that the proposition is somehow imprecise, identifying the source of information either specifically or by implication and marking limitations of the information or identifying the perspective from which the proposition is true. Secondly, attitudinal stance conveys the speaker's attitudes, feelings, or value judgments. Lastly, style stance describes the manner in which the information is being presented (ibid). "Grammatical Realization" (Hunston, 2000:60) is a common descriptive parameter for stance adverbials. The term explains that these realisations might appear "as a single adverb, adverb phrase, noun phrase, prepositional phrase, finite subordinate clause, or non-finite subordinate clause" (ibid:60). Most adverbials can occur freely in different clause positions: initial, pre-verbial, post-verbial and final (ibid:61). Conrad and Biber (2000) list the following examples of epistemic stance:

1. Maybe she put grease in it. (CONV)
2. I'm sure I've spoken to him. (CONV) (Conrad and Biber, 2000:58)

In the first example, the speaker uses the adverb "*maybe*" to convey doubt about the following proposition: *she put grease in it*. Similarly, the adjective *I'm sure* is used in the second example to highlight the speaker's certainty about the following proposition *I've spoken to him*. There are three major parameters for the adverbial stance markers, as the two linguists Conrad and Biber elaborate. Firstly, there is the semantic class. Secondly, there is the grammatical realisation. Lastly, there is the placement in the clause. (ibid:58)

As such, expressions of stance are a general category in linguistics, realised by many grammatical forms: the main focus in this research is the grammatical form of adverbs, particularly attitudinal and epistemic stance adverbials.

2.4 Stance across other frameworks

Stance expressions are among the most active research topics in both general linguistics and corpus linguistics. There are three widely cited frameworks that will be analysed in this section, showing how speakers and writers position themselves to their audiences: Hyland's metadiscourse (2019), Martin and White's appraisal system in English (2005), and Halliday and Matthiessen's

theory of attitude (2014), defined under Systemic Functional Grammar (SFG). Their importance lies not only in their invaluable theoretical contributions but also in their points of agreement and disagreement regarding the nature of the stance itself and the methodological requirements for its systematic study and analysis.

All three frameworks share the interpersonal dimension of language. Whether the stance is labelled modality, appraisal, or metadiscourse. Each framework contributes to understanding how a writer/speaker shapes their stance for their reader/listener. Halliday (2014) describes “modality” as “the speaker’s judgement, or request of the judgement of the listener, on the status of what is being said” (Halliday, 2014:172): a definition that is interactive and relational. Similarly, In their book, J.R. Martin and P.R.R White (2005) define appraisal as “how writers/speakers approve and disapprove, enthuse and abhor, applaud and criticize, and with how they position their readers/listeners to do likewise” (2005:1). These interpersonal relational dimensions are, once more, foregrounded and asserted in Hyland’s (2019) definition of metadiscourse as “the cover term for the self-reflective expressions used to negotiate interactional meanings in a text, assisting the writer (or speaker) to express a viewpoint and engage with readers as members of a particular community” (Hyland, 2019:43,44). Hence, according to all three frameworks, stance is not apprehended as a mere concrete cognitive unit in the language, or solely as an expression serving a clause’s proposition, but rather as a social communicative tool that relies on the interpersonal dimension of language. In other words, stance is a positioning act of a speaker/writer that describes their epistemic, attitudinal, or stylistic stance to their audience interpersonally. This shared interpersonal aspect is significant because it reveals that the three frameworks are not defining completely different phenomena, but rather distinct analytical perspectives applied to the same underlying phenomenon; each tackles a certain layer while sharing a common real communicative aspect.

The disagreements among the three frameworks begin, however, with the analytical concepts and with how each scholar links stance expressions to grammatical forms. Halliday’s (2014) SFG distributes stance-like meanings across several subsystems within the interpersonal metafunction. Modality operates as a moderate concept between positive and negative polarity, which he defines as the “opposition between positive (*It is. Do that!*) and negative (*It isn’t. Don’t do that!*)” (Halliday, 2014:172). Crucially, Halliday extends these resources beyond the grammatical core system of modality to incorporate broader interpersonal assessments. These assessments “relate to either to the proposition being exchanged (e.g. *Fortunately it is*: it is, which is fortunate) or to the act of exchanging it (e.g. *Frankly it is*: I am telling you frankly it is)” (ibid:173). These items function as comment adjuncts that provide evaluative commentary from the speaker and thus realise what corpus linguistics would classify as epistemic or attitudinal stance. However, the term stance does not appear. Instead, stance-like meanings are distributed across modality, modal assessment, and interpersonal

adjuncts. However, the critical aspect of Halliday's (2005) modality is that, within SFG, these resources are distributed across modality, modal assessment, and interpersonal adjuncts as separate subsystems, without being classified under a single taxonomic categorisation. Using this framework, an analyst would therefore need to reconstruct stance from multiple subsystems realised through different grammatical resources, modal devices, and interpersonal adjuncts. As such, this methodology complicates stance systemic analysis, quantification, and retrieval, and as a consequence, the risk of analytical inconsistency increases.

Martin and White's (2005) appraisal theory framework represents a significant theoretical analysis of the interpersonal aspect and shares with Halliday's (2014) SFG an important question that relates to how the evaluative meaning is constructed across texts and genres. Meaning in Appraisal theory is based on three notions: ideational resources that create the experience: "who is doing what to whom, where, when, why and the logical relation of the flow", interpersonal resources that reflect how people are interacting, including the feelings they try to share, and textual resources that are viewed as an overlapping area between ideational and interpersonal resources. They are about the information flow and how ideational and interpersonal meanings are distributed (Martin and White, 2005:7). Martin and White (2005) classify appraisal into three main categories: Attitude, Engagement, and Graduation; together, they establish the meaning of appraisal. While Attitude captures our feelings, including emotional reactions and judgments, Engagement is concerned with "sourcing attitudes and the play of voices around opinions in discourse" (ibid:35). Graduation is concerned with "gradability" of feelings: how weak or strong the feeling is (ibid:37). Martin and White (2005) highlight a very interesting note that "appraisal is probably most closely related to the concept of stance, as developed by Biber and his colleagues in their corpus-based quantitative studies" (ibid:40). This acknowledgement is theoretically significant because it denotes that this framework and Biber's are not two competing theories, but they are rather two different lenses, each of which helps view stance at different levels of abstraction. However, Appraisal theory's main concern is discourse semantics: how the evaluative meaning is constructed through the three main categories of Appraisal: attitude, engagement, and graduation within a given discourse, rather than identifying stable, lexical, grammatical and functional linguistic items. Such a holistic approach, despite being theoretically rich, is less well-suited to large-corpora analysis, as there will always be a high rate of inconsistency among analysts during manual data processing, as the concept is scattered across many categories and sub-categories, which makes the analysts' tasks very challenging and more vulnerable to error.

By contrast, Hyland's (2019) metadiscourse model offers a more explicit classification. He groups stance expressions by their discourse function, distinguishing hedges, which "withhold commitment and open dialogues, e.g., '*might; perhaps*', from "boosters", which "emphasise certainty

or close dialogue”, e.g., “*in fact, definitely*”, from “attitude markers”, which are used to express “writer’s attitude to proposition”. (Halliday, 2019: 58). Such a functional classification has descriptive methodological power and, with precision, highlights the communicative reality of various discursive contexts. The metadiscourse model consistently aligns with its theoretical foundation, which defines it as “self-reflective expressions used to negotiate interactional meanings ... assisting the writer or speaker to express a viewpoint and engage with readers as members of a particular community” (Hyland, 2019:43,44). That means the classification prioritises the communicative roles of stance over the grammatical realisation. Nonetheless, this functional classification raises some challenges for analysts processing corpus-based studies, specifically large corpora, because the analysis would rely on discourse function rather than on grammatical form. Hyland’s (2019) framework permits the same linguistic device to shift between metadiscourse and the propositional content depending on the context. As Hyland explains, metadiscourse:

“is a relative concept in that text items only function as metadiscourse in relation to another part of the text. So what might be metadiscourse in one rhetorical context may be expressing propositional material in another, and analysts must always examine each item individually to determine its function” (ibid.:28).

As such, this context-sensitivity, while theoretically comprehensible, means that the framework does not yield any stable, form-based categories that can be identified, extracted, and quantified across large corpora without interpretive intervention by the analysts. This would, of course, lead to more inconsistent results.

Comparing the three frameworks reveals a fundamental disagreement running through stance research and analysis as a whole: the theoretical richness of the frameworks and the methodological processing. Halliday’s (2014) modality offers a comprehensive view of the grammatical resources through which stance is realised, grounded in a holistic theory of language as a social system. Martin and White’s (2005) model of appraisal represents a semantic overview of how evaluative meaning is constructed. Hyland’s (2019) metadiscourse model demonstrates how stance expressions interact communicatively in discursive contexts. However, all three frameworks, surprisingly, prioritise a functional or semantic approach over grammatical stability, which is required for consistent analytical procedures. None of these frameworks could provide a unified taxonomic reference for stance expressions that analysts can systematically use to spot, retrieve, quantify, and compare registers diachronically without a significant burden of deep manual classifying, testing, and interpreting.

Given the above comparison, it is clear that Biber et al.’s (1999) framework stands apart. Unlike the three frameworks analysed, Biber et al. (1999) grounded their stance in lexicogrammatical features, e.g. modals and adverbials, which are form-based and flexible to extraction and analytical procedures across any given corpus of any given size. As such, Biber’s framework offers distinct analytical tools and priorities for different sets of research questions. The present study’s main focus

on stance adverbials in the BNC1994 and BNC2014 requires these features. Biber's framework is therefore adopted to enable extensive analysis of stance adverbials, aiming for cohesive and consistent results in diachronic studies of large corpora. Thus, the choice of Biber's framework (1999) does not mean the rejection of other frameworks. In fact, the sophistication of Hyland (2019), Martin and White (2005), and Halliday's (2014) studies stands out theoretically as more abstract frameworks that explain the multi-layered nature of stance expressions in linguistics, yet they are less suited to corpus-based analysis of large corpora.

2.5 Previous Research

Research on stance has become a central area in corpus linguistics because it denotes how speakers and writers in a given discourse position themselves toward propositions and signal degrees of certainty, evaluation, and commitment. Whether the research is based on narrowly focused sets of lexical items or moves toward broader concepts and their applications, they all highlight limitations that point directly to the need for the present study in corpus linguistics. This section addresses the first axis by reviewing four stance-based studies that process stance with a corpus-based approach, focusing on Biber's lexicogrammatical categorisation.

Three of the four studies examined in this section, Banaitytė (2017), Ağçam (2015), and Daughalich and Han (2024), draw on Biber's functional framework on stance either partially or fully, combining it with other complementary frameworks. Whereas Simon-Vandenberg (2008) ground in Paradis (2008) model of scale and boundedness. They all employ corpus-based quantitative methods to investigate the semantic fields, frequencies, and distribution of stance adverbials across the selected corpora. Daughalich and Han's (2024) and Ağçam's (2015) studies additionally stance in English as a second language learners. However, only Daughalich and Han's (2024) use a qualitative approach, extending the research beyond the quantitative approach. Despite various grounding in various frameworks, directly or indirectly related to Biber's, and that the four of them are based on straightforward analysis of expressions of stance, extracting the frequencies and distributional properties of stance adverbials, each of them differs in its analytical perspective, register coverage and size, and the depth of its analysis of stance as a multilayered linguistic device.

The first visible point of agreement among the four studies is their grounding in Biber's framework of stance as the primary lens for understanding its dimensions. Simon-Vandenberg (2008) cites for Biber et al. (1999), however, using a complementary framework of scale and boundedness by Caritas Paradis (2008). Within this model, Paradis proposes that linguistic meanings might be transformed into scalar dynamic forms rather than static categorical ones. From this perspective, combinations of modifiers and stance adverbs represent dynamic constructions that highly affect the degree of epistemicity of a stance. The corpus data of this study are drawn from the BNC, examining the behaviour of the certainty adverbs *certainly* and *definitely* when controlled by

degree modifiers *almost* and *most*, respectively. Simon-Vandenberg (2008) argues that each adverb is differently conceptualised. While *certainly* is conceptualised as a totality word, *definitely* is conceptualised as a scalar word, and its function is to convey subjective emotional commitment. Banaitytė (2017) draws directly on Biber et al.'s framework, similarly drawing on Martin and White's (2005) Appraisal theory to spot the interpersonal and ideological dimensions of stance in journalistic discourse. Ağçam (2015) and Daughalich and Han (2024) based their studies explicitly on Biber's framework, with Ağçam (2015) drawing on Biber (2006), while Daughalich and Han (2024) combine Biber and Finegan (1989), Biber et al.'s (1999) stance framework and Hyland's (2005) model. As such, this shared theoretical background allows further comparisons across the studies, highlighting that Biber's taxonomic categorisation is flexible and can be adapted by many analysts for various research objectives, thereby assuring the versatility of investigations, from narrow, limited lexical sets to broadly general ones.

There is more than one aspect of divergence among the four studies. Perhaps the most apparent one is the corpus variation across the studies. Simon-Vandenberg (2008) adapted to a very limited lexical set, examining two adverbs of certainty when combined with degree modifiers to investigate scalar applicability and the degree of epistemicity they convey. The researcher extracted 500 instances from the BNC for each adverb: *certainly* and *definitely*. While this narrow focus enables a very high degree of precision, it hinders generalizability to the language as a whole. The researcher could demonstrate that adding intensity modifiers to adverbs, specifically *almost certainly* and *most definitely*, transforms the former into a conceptualised totality word, while the latter is conceptualised as a scalar word, aligning with Paradis's (2008) boundedness and scale framework. On the other hand, Banaitytė (2017) used a wider-scale corpus. The researcher used a compiled corpus of 50,000 words, extracted from two British newspapers, specifically opinion articles, investigating media discourse across the two newspapers. Such a corpus, considered larger than Simon-Vandenberg's (2008), still constrains the research scope and consequently narrows the analytical lens and the diversity of the findings, raising once more the limitation of the findings' generalizability. Daughalich and Han (2024) compiled two balanced corpora: a learner corpus consisting of 804,935 words, and a reference corpus consisting of 802,490 words, examining certainty adverbs, e.g. *actually*, *clearly*, *in fact*, *obviously*, and *really*. Lastly, Ağçam's (2015) is the broadest corpus among the four studies. The researcher compiled four sub-corpora consisting of approximately 2 million words, based on a predefined list of 47 attitudinal stance markers explicitly mentioned in Biber's (2006) framework of stance, including adverbials such as *surprisingly*, *unfortunately*, *hopefully*, and *ironically*.

As highlighted above, the limitedness of the examined lexical sets and the small corpus size, as in Simon-Vandenberg (2008) and Banaitytė (2017), hinder the generalizability of the findings, making the study very limited and the results very specific to the examined sets. Furthermore,

although three of the four studies are grounded in Biber's (2006) and/or Biber et al.'s (1999) framework, some limitations remain, underscoring the importance of the current study. Simon-Vandenberg (2008) and Daughalich and Han (2024) used certainty adverbs as epistemic markers, without considering attitudinal stance. Banaityté (2017) used both epistemic and attitudinal stance markers within a very limited corpus size. Whereas Ağçam (2015) based her research solely on attitudinal markers, without considering epistemic stance markers.

Another axis of variation among the studies concerns the registers under investigation in each of them. While Banaityté (2017) examined argumentative journalistic discourse in two British newspapers, namely, The Guardian and The Times, Simon-Vandenberg's (2008) corpus was selected from the BNC from multiple registers, centring her research on the semantic aspects and the effect on functional dimensions of two modified certainty adverbs, rather than on their register-specific distribution. Ağçam's (2015) and Daughalich and Han's (2024) studies both operate on academic discourse. While Daughalich and Han (2024) focus on comparing Chinese MA theses with internationally published research articles to extract the use of specific certainty adverbs, showing their epistemicity, Ağçam's (2015) focus is on doctoral dissertations produced by native speakers and non-native speakers of English. The effect of choosing such corpora is a body of research more aligned with academic and news registers, with very little attention to the other registers, which are rich in stance expressions, such as conversation, fiction, e-language, etc. This imbalance is significant because even Biber's corpus-based findings demonstrate that functional and distributional properties of stance expressions vary considerably across registers.

The four studies also vary in their methodologies, particularly in the qualitative and quantitative analyses of each of them. Simon-Vandenberg (2008), Banaityté (2017), and Ağçam (2015) all analyse the frequency of stance adverbials in the specified registers using normalisation and conduct the necessary statistical tests to assess significance and relevance. It is important to note, however, that the quantitative analysis alone cannot provide a full image of the use and exploration of stance expression. This can be conducted only via manual analysis. This exact limitation was addressed by Daughalich and Han (2024) in their study: they not only revealed that students underuse stance expressions via quantitative analysis, a shared aspect with other studies, but also that they misuse them pragmatically and syntactically. As the study found that Chinese students overuse *actually*, which witnesses an influence from the Chinese language, pairing it with *actually speaking*, they pair *really* with abstract rather than concrete referents, and they collocate *clearly* with the pronoun *it* rather than with textual references that disciplinary conventions require. Such findings are impossible to concretise without a proper qualitative analysis. Together, the four studies suggest that a comprehensive study of stance must include both qualitative and quantitative analyses, as relying solely on quantitative analysis would not yield consistent findings.

The synchronicity of the four studies is yet another limitation. Simon-Vandenberg (2008), Banaitytė (2017), Ağçam (2015), and Daughalich and Han (2024) all examine a specific scope of stance adverbials at a particular point in time, focusing on distinct corpora of a certain size that do not demonstrate any diachronic perspective. Ağçam (2015) extracted the doctoral theses of native and non-native speakers of English between 2005 and 2012; however, the study was not conducted diachronically. Simon-Vandenberg's (2008) and Banaitytė's (2017) studies are concerned with the corpora extracted from the BNC and the newspapers, respectively, without constraining their studies to any time range. Furthermore, Daughalich and Han (2024) do not mention the temporal dimension of their studies. Only Ağçam (2015) asserts a temporal dimension of the data, without using them diachronically. All the studies neglect the diachronic dimension, and by contrast, demonstrate the effect or the change synchronically. This is particularly a significant gap, given that the BNC facilitates the analytical tasks by providing corpora that belong to two different periods, making such an investigation possible. This unprovided aspect, shared by all four studies, prevents the reader from determining whether the stance expressions exhibit specific patterns over time. Such absence of the diachronic dimension makes the studies' contribution less acknowledged by analysts interested in extracting the patterns of use of expressions of stance, which might show their evolution and change over time.

Collectively, the four studies provide a productive yet incomplete empirical foundation for corpus-based studies of stance adverbials. The four of them are grounded in Biber's lexicogrammatical framework of stance across various registers within research contexts. Each of them is concerned with a specific layer of stance expressions, which are known for their multilayered nature. Furthermore, three studies use a quantitative research method, whereas only one employs both quantitative and qualitative approaches. The four studies leave three unaddressed gaps: a restricted lexical scope that limits generalizability, a narrow register coverage that does not fully reflect the distributional range of stance adverbials, and a very strict synchronic processing of data, leaving the diachronicity of the studies an unanswered question. Contrasting the limitations of these studies with the current thesis demonstrates why it stands out as a diachronic research paper on expressions of stance that uses an analytical lens showing two layers of stance: frequency layer, categorisation as epistemic or attitudinal based on the interpretation, not solely on the framework's classification, applying both qualitative and quantitative approaches. Furthermore, the study's findings are linked to subjectification theory (Traugott 1989) and Halliday's (2014) register theory, showing different trends in language over 20 years, relying on the BNC corpora available for two time periods.

Moving to the second axis of the literature review, which examines some studies based on data from the British National Corpus. As one of the most reliable sources for corpus-based studies, the BNC, with its two versions from 1994 and 2014, is an invaluable resource for language

researchers seeking to trace developments and evolution in language, both synchronically and diachronically. This includes changes in frequency distribution, semantic functions, and discourse over approximately twenty years. The four studies examined in this section are Laws et al. (2017), Meng (2019), Kaatari and Larsson (2019), and Aijmer (2020), all of which draw on BNC data, either directly or in combination with other corpora, to investigate linguistic changes. Nonetheless, despite sharing the same resource and diachronic approach, these studies differ fundamentally in the linguistic phenomena examined, their analytical focus, methodological approach, and the findings they reveal, highlighting the breadth of what BNC-based research can elucidate and the specific gaps it has to address. Each of these differentiation points reveals what BNC-based research has contributed and which gaps remain to be addressed, highlighting the importance of this thesis.

A central point of agreement among three of the four studies is that they share the concept of grammaticalization as the explanatory framework for the diachronic changes that take place. However, the researchers approach this concept from different theoretical perspectives and with varying degrees of explicitness. Kaatari and Larsson (2019) ground their analysis directly in grammaticalization theory, drawing on Thompson and Mulac (1991) and Traugott and Heine (1991), who frame linguistic grammaticalization as the process through which constructions develop from main-clause uses to parenthetical and epistemic functions. Kaatari and Larsson (2019) examine “*I think*” and “*I’m sure*” meanings via two dimensions: syntagmatic and paradigmatic, according to Thompson and Mulac (1991), extracting their data from BNC1994 and BNC2014. The syntagmatic variability refers to the positional flexibility of a construction within a clause. As such, it is thus concerned with the ability of *I think*, and *I’m sure* to occur in positions other than clause-initial position. On the other hand, the paradigmatic variability is concerned with the extent to which that is omitted in clause-initial position. Their findings demonstrate that both constructions exhibit increased positional flexibility and higher rates of *that* omission diachronically. Kaatari and Larsson (2019) confirm that their findings align with grammaticalization, especially with the increased *that* omission, and the positionality of these constructions shows diversity in initial, medial, or final clausal position, considering this as a sign of grammaticalization. Meng (2019) similarly processed the concept of grammaticalization, drawing on Hopper and Traugott (1993). The study demonstrates that *like* and its variants have developed diachronically and that they align with grammaticalization theory, which is a gradual, slow process of constructions, involving semantic and textual changes.

Using BNC data alongside COHA covering two centuries, Meng (2019) documented the progress witnessed for *like*, indicating various meanings along its historical change: *similar to* from the 1300s, *as if* in 1386, *approximately* in 1590, and *for example* in 1886. A more recent functional change of *like* as a discourse-pragmatic marker for focus in 1959 and as a quotative marker in the 1960s onwards. Such a trend is a clear example of grammaticalization. Aijmer (2020) extends this

shared theoretical perspective to the lexical-pragmatic domain, tracing the change that *absolutely* underwent from a degree modifier to an emphasiser and, finally, to a discourse marker. Drawing on Traugott's (2003) theory of grammaticalization and Himmelmann's (2004) grammaticalization framework, which is characterised by three types of expansion: host-class, syntactic, and semantic-pragmatic expansion. The findings exhibit that *absolutely* demonstrates a significant decline in the pre-adjectival use, alongside a significant rise in discourse-marker function, and that it extends to new semantic-pragmatic contexts. suggesting that *absolutely* underwent a redistribution across functional contexts. Laws et al. (2017) is the only study that does not invoke grammaticalization as a theoretical framework, focusing instead on derivational morphology in verb formation, drawing on the morphological productivity theory of Bauer (1983, 2013), Plag (1999, 2004), Biber et al. (1999) and Marchand (1969) to examine lexical diversity, density, and creativity. Their findings demonstrate that the suffixes *-ize*, *-ify*, and *-ate* have increased in token frequency while the suffix *-en* has declined; as such, the findings show consistency with the broader pattern documented by the other studies' interest in grammaticalization, whereby gradual changes over time affect the function of linguistic constructions. The studies collectively show that grammaticalization has become an explanatory framework for linguistic diachronic changes, despite their differences in depth and in the processing of various frameworks.

An important point of disagreement among the studies is the linguistic phenomena under investigation and the methodologies used to construct their corpora. Kaatri and Larsson (2019) work on the syntactic aspect, restricting their analysis to two epistemic constructions, *I think*, and *I'm sure*, using comparable data from BNC1994 and the spoken BNC2014, totalling 4,233,962 and 11,422,617 million words respectively. High precision is primarily the advantage of a narrow lexical scope for analysis, but it limits the generalisability of the findings to other epistemic forms. Similarly, Aijmer (2020) focuses on a single lexical item, *absolutely*, extracting concordance lines from the BNC1994 and BNC2014, totalling approximately 4,788,882 and 5,014,655 million words, respectively. However, she expands her research through diverse syntactic forms, e.g. adjectival, verbal, prepositional phrases, discourse markers, and response markers, highlighting the full functional range of the construction. This design enabled the researcher to document an increase in frequency and to highlight syntactic, semantic and pragmatic changes, suggesting an on-going grammaticalization process. Laws et al. (2017) examine a limited but broader set of four verb-forming suffixations: *-ize*, *-ify*, *-ate*, *-en* across both BNC corpora, applying log likelihood tests with Bonferroni correction and outlier replacement procedures to control the contribution of individual speakers. Such a high level of statistical accuracy is unmatched across the other studies reviewed in this section. Laws et al.'s (2017) analysis remains at the level of token frequency distribution, with limited attention to the contexts in which some suffixes are prioritised over others, a gap that a qualitative analysis lens could

easily address. Meng (2019) additionally uses the Corpus of Historical American English (COHA) alongside the BNC to track *like* and its variants over a two-century span, from the 1810s to the 2000s. While such a design is capable of capturing the grammaticalization phenomenon diachronically over a long period of time, contrasted with the small 20-year window the BNC allows, it also introduces a comparability issue that is not addressed by Meng's (2019) study: COHA is an American English Corpus, whereas the other studies investigate British English, and the conflation of the two varieties limits the degree to which Meng's (2019) findings can be directly incorporated with other BNC-based research.

The four studies differ in their methodological approach, in particular, the balance achieved between the qualitative and quantitative depths of each study. Laws et al. (2017) adopt the most explicitly quantitative approach, with the strongest statistical analysis among the studies concerned. Their study focuses on normalised token frequencies, log-likelihood-based significance testing with a Bonferroni correction to control for multiple comparisons, and a procedure for replacing outliers to ensure that significant findings are not driven by a small number of highly productive individual speakers. Such a high standard of precision is unmatched among the other three studies. Laws et al.'s (2017) study is based mainly on frequency counting involving other elements like age and gender. Kaatari and Larsson (2019) adopt a quantitative comparison of overall frequency combined with manual analysis of two categorical variables: clausal positioning and *that* omission, across more than five thousand tokens, yielding precise data while operating within a distributional paradigm. The strength of this study design lies in its ability to track precise syntactic changes across the twenty-year interval, whereas its limitation is that it does not engage with the functional or pragmatic dimensions of epistemic stance markers beyond the syntactic positioning of the forms investigated. Aijmer (2020) integrates quantitative frequency analysis with a qualitative examination of the syntactic and semantic constructions in which *absolutely* occurs, including age and gender elements, stylistic features such as hyperbole, and discourse-pragmatic functions. Such a combination of methods ensures the precision of the frequency results and enables a profound functional analysis, highlighting additional aspects of the phenomenon under investigation. By Contrast, Meng's (2019) study is the least robust, as the researcher used Sketch Engine to extract collocational and frequency data from the BNC and COHA and presents concordance lines, illustrating different historical meanings of *like* and its variants. The analysis remains largely descriptive, documenting frequency shifts across two centuries without statistical testing for significance, systematic functional processing, or engagement with further theoretical implications of the observed phenomena beyond grammaticalization.

Considered collectively, these four studies provide a productive yet incomplete empirical foundation for corpus-based investigations of language change and evolution in English. They are

grounded in grammaticalization as the dominant explanatory framework, and they mostly explain the rise in frequency in terms of functional redistribution. They diverge, however, in the linguistic phenomena examined, the comparability of corpora and their size, the depth of their engagement with social variables, e.g. age and gender, and the dimension of their methodological processing. As such, the literature leaves two gaps that the present thesis addresses. First, despite their centrality to the types of grammaticalization and delexicalization, stance adverbials are not the focus of any of them, whereby Kaatri and Larsson (2019) document epistemic constructions, Aijmer's (2020) engages with intensifiers, and Meng's (2019) study proposed discourse markers. Laws et al. (2017) investigated verb formation suffixes. All studies except Laws et al.'s (2017) carry stance meanings, but the relationship between grammaticalization and stance has not been fully explored by any of them: the epistemic and attitudinal markers of stance adverbials remain an unexplored area of research within the three studies. Second, the methodological imbalance between the three studies that prioritise a quantitative approach and statistical rigour, as in Laws et al. (2017) and those that prioritise qualitative depth and functional interpretation, as in Aijmer's (2020), with one study, however, Meng's (2019), provides descriptive frequency trends without statistical testing or deep functional analysis. As noted, balanced research combines qualitative and quantitative approaches to meaningfully explain observed patterns. By contrast, the current study provides qualitative and quantitative analysis of stance adverbials in British English, drawing its data from the BNC1994 and the BNC2014, working on semantic analysis to fully understand and categorise the data under investigation diachronically to trace the attitudinal and epistemic expressions of stance adverbials and their evolution over a span of twenty years, and links the findings to Traugott's subjectification theory (1989), so that subjectivity is rather frosted as a deep speaker's or writer's orientation in academic and e-language registers of the BNC. As far as the studies are concerned, none addressed this within their topicality and made it further explicit in their work.

Lastly, the third dimension addressed in the literature review section is the diachronic studies, which range from a decade to a century and a half. The four studies reviewed in this section are Poole et al. (2019), Kuhi and Rezaei (2020), Abdeljaoued (2026), and Tantucci (2015). They all share a fundamental commitment to tracing how writers' evaluation has changed diachronically; however, they focus on different genres, anchor their studies in different theoretical frameworks, and use different corpora in scope and density. An interesting point of convergence among three of the four studies is their agreement on the concept of *factualization*. By this, Tantucci (2015) means the shift from hedges and tentative epistemic positioning to a more assertive, confident, and overtly subjective stance-taking. Despite some profound differences, this agreement on a single trend suggests that the *factualization* might constitute a general feature of academic discourse development across the modern period, though how far the corpora are heterogeneous, combined with methodological

perspective and the analytical depth of each study, reveals some unaddressed gaps in the literature as a whole. Only Abdeljaoued's (2026) findings do not align with *factualization*.

An important point of agreement among the studies is how each of them highlights that epistemic markers have shifted diachronically from uncertainty towards certainty and assertion. Tantucci (2015) highlights *factualization* as a cornerstone in his study, describing the process through which epistemic predicates originally conveying weak certainty, e.g. *I think*, *I believe* and *I reckon*, develop more factual meanings over time. Drawing on Langacker's (1991, 2008, 2009) notion of the epistemic control cycle, Tantucci (2015) demonstrates through the analysis of the BNC extracted data that *I think* can convey inclination, action, or result depending on context. However, due to its polysemies, *I think* is a candidate to undergo the process of factualization. Tantucci (2015) extends this framework diachronically by analysing *Io penso*, the Italian equivalent of *I think*, using the diaCoris corpus, showing that the modern use of *Io penso* in Italian is more factualized than it was 150 years ago. Kuhl and Rezaei (2020) demonstrate an equivalent trend in applied linguistics discussion sections, finding that hedges such as *significantly* across the periods from 1996 to 2006 and from 2006 to 2016, while self-mentions increased. Abdeljaoued (2026) extends this finding to UK universities' doctoral theses, demonstrating that boosters declined slightly, whereas booster verbs and adverbs increased from 1950 to 2016. Attitudinal and authorial stance markers' frequency increased. Poole et al. (2019) process their study within the specialised domain of chemistry and specifically the studies that cite the 1972 seminal study entitled "Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking". Their findings show that non-modal hedging declined modestly, while boosters increased over 45 years. Additionally, the epistemic markers encoding certainty increased, while those reflecting doubt and uncertainty decreased. The studies, therefore, provide strong evidence that academic writing has become more assertive epistemically over recent decades. However, each study differs in how it methodologically addresses this change.

The first point of disagreement among the studies is their corpus design. Poole et al. (2019) collected the most tightly focused corpus from 328 open-access articles on chemistry and chemotaxis published from 1972 to 2017, totalling approximately 2,408,337 million words. Such a single-topic approach in design allows the authors to control for epistemic stance within a stable discourse register diachronically, raising once more in the literature review the question of generalisability. Similarly, Kuhl and Rezaei (2020) adopt a narrow, discipline-oriented approach, analysing discussion sections from three applied linguistics journals (*Applied Linguistics Journal*, *English for Specific Purposes Journal*, *Modern Language Journal*) across two decades, yielding three subcorpora totalling 914,679 words. While this design can highlight the changes that linguistic phenomena undergo, it restricts the analysis to the discussion sections, excluding other sections such as introductions, methods, results, and the abstract. Abdeljaoued's (2026) study has a larger disciplinary

dimension as it examines 240 UK doctoral theses across eight disciplines: Biology, Chemistry, Engineering, Physics, History, Linguistics, Philosophy, Sociology, over a period of 66 years from 1950 to 2016, resulting in a large corpus of 13 million words of cleaned text. This approach enables comparability between the hard sciences and humanities, allowing testing of whether stance evolution differs across branches. However, grouping these theses into disciplinary categories might obscure subdisciplinary variation. Furthermore, restricting the study to UK universities' theses only limits its generalisability to other academic cultures. Tantucci's (2015) study examines three epistemic verbs, *I think*, *I believe*, and *I reckon*, across the entire British National Corpus and expands to the Italian language using diaCoris spanning 1861 to 2001. While this is considered the widest temporal range among all studies, the shift from English to Italian introduces a comparability problem. The findings about *factualization* in Italian might not translate directly into English, given the socio-historical and epistemic marking differences between the two languages. As such, each corpus design enjoys some advantages, yet highlights some limitations: Poole et al. (2019) and Abdeljaoued (2026) show disciplinary specificity and temporal depth. Tantucci's (2015) research shows cross-linguistic comparability, whereas genre consistency and representational breadth are manifested in Kuhi and Rezaei (2020) and Abdeljaoued (2026).

The four studies draw on different frameworks to explain the diachronic change in their data. Tantucci (2015) grounds in cognitive-functional linguistics, drawing on Traugott's (1989, 1995, 2003) subjectification framework. Tantucci (2015) argues that *factualization* is a process in which constructions acquire speaker-oriented and more subjective meaning over time. As such, Tantucci's (2015) *factualization* is consistent with grammaticalization theory, which predicts that linguistic forms move from objective, propositional content toward subjective forms and discourse-pragmatic functions. The epistemic control cycle framework by Langacker (1991, 2008, 2009) provides the conceptual dimension for observing this change. Kuhi and Rezaei (2020) adopt Hyland's (2005b) model of metadiscourse to identify and classify stance markers. However, the authors do not expand the research to explain why the language phenomenon underwent a specific change. Abdeljaoued's (2026) study draws on Hyland's (2005a,b) stance model to examine the use of hedges, boosters, and self-mention. Additionally, Abdeljaoued (2026) draws on Becher and Trowler's (2001) typological scheme of disciplinary knowledge domains. The allocation of the theses to the mentioned categories was guided by their rhetorical and methodological specificities rather than by institutional or departmental orientations; this approach highlights the study's precision. This pragmatic procedure allows for an epistemic breadth, excluding any deterministic disciplinary boundaries. Abdeljaoued (2026) treats each thesis as a whole study, not including certain sections and excluding others, as researchers believe that the disciplinary variation in rhetorical organisation makes section-based sampling less reliable. Poole et al.'s (2019) study draws on both Hyland (2005a,b) and Biber (2006)

for stance identification and categorisation, analysing 8 central modal auxiliaries, 77 hedging devices, and 54 boosters. Their findings align with Tantucci's (2015) regarding *factualization*: a decrease in the use of modal auxiliaries, non-modal hedges, and stance markers indexing doubt and uncertainty, while boosters and epistemic stance markers of greater certainty increased.

Taken together, the four studies reveal both convergence and disagreements in the dimension of diachronic studies. On the one hand, there is an evidence for the shifts of the expressions of stance over time, including declining reliance on hedging devices in highly specialized domains, as shown in Poole et al. (2019) study, relative stability in overall stance density, as demonstrated in Abdeljaoued (2026), increasing visibility of authorial voice through self-mention, a trend observed in both Abdeljaoued (2026) and Kuhi and Rezaei (2020), and increasing number of epistemic forms that develop a factual meaning conveying a subjected form of certainty as Tantucci's (2015) study demonstrates his notion of *factualization*. On the other hand, these findings are shaped by varying methodological choices, including the size and range of the corpus, ranging from a highly controlled and specialised dataset in Poole et al. (2019), to a larger dataset with a more relaxed control in Abdeljaoued (2026) and Kuhi and Rezaei (2020), reaching the broad resources of the BNC and diaCoris corpora in Tantucci's (2015) study that extends investigating linguistic constructions over 150 years. The examined datasets widely differ among the studies, from examining very narrow lexical sets as in Tantucci's (2015) *I think, I believe, and I reckon*, in addition to the Italian *Io penso*, to wider investigations of hedges, boosters, attitude markers and self mentions as in Abdeljaoued (2026) and Kuhi and Rezaei (2020). Whereas Poole et al. (2019) examine modal auxiliaries, hedging devices, and boosters. The framework followed for stance categorisation also varies among the four studies: while Abdeljaoued (2026) and Kuhi and Rezaei (2020) draw primarily on Hyland's (2005) model of metadiscourse, Poole et al. (2019) use the same model combined with Biber's (2006) taxonomy of stance. Whereas Tantucci (2015) draws on Traugott's (1989, 1995, 2003) subjectification framework, combined with an analytical perspective grounded in Langacker (1991, 2008, 2009). Despite the valuable contributions of the four studies, they all share limitations that underscore the importance of the current study. Primarily, the scope of each study is not stance adverbials as a basic component of an expression of stance, but the studies explored other elements as boosters, hedging devices or modal auxiliaries. The semantic layer of stance has not been explored in any of the studies. Such limitations provide motivation for the present study, which aims to integrate frequency-based analysis and comparison together with detailed semantic classification, focusing particularly on stance adverbials across registers diachronically.

It is noteworthy that the AP1994 register was previously examined by Biber et al. in their invaluable resource, *Written and Spoken Grammar of English*, in 1999, where they introduced some interesting findings about the academic prose register of the BNC1994. They noticed that stance

adverbials are more common in academic prose than in other registers, such as fiction and news, and are the most frequent category among other types, including circumstance and style adverbials. However, academic prose consists of attitudinal expressions of stance more than other registers. They also document that manuals, textbooks, and technical reports include the author's attitude toward the material and sometimes comments on what is expected or surprising (Biber et al., 1999:859-861).

3. Data and Methods

3.1 Theoretical framework

This MA thesis is grounded in the framework of stance expressions by Biber and Finegan (1989), Biber et al. (1999) and Biber (2006). As per the framework, epistemic adverbials express “the attitude of the speaker/writer with respect to the proposition contained in the matrix clause, often conveying certainty or likelihood” (Biber, 2006:88). Attitudinal adverbials, meanwhile, “tell of the writer's or speaker's attitude toward the proposition, typically conveying an evaluation, value judgment, or assessment of expectations” (Biber et al., 1999:856). However, both stance types share an important aspect: “epistemic stance adverbials and attitude stance adverbials both comment on the content of a proposition” (ibid:854). It is important to note that the framework exclusively names certain adverbs, categorising them as attitudinal or epistemic stance adverbials. The table in Appendix 5 explores each type of stance: epistemic or attitudinal, and their relevant adverbs exclusively identified by Biber and Finegan (1988), Biber et al. (1999) and Biber (2006).

In Biber et al.'s (1999) framework, stance adverbials are classified under the grammatical/lexico-grammatical category of stance expressions. Stance adverbs have a separate category that distinguishes them. There is a single general classification of adverbial categories. This category of adverbials is further classified into circumstance adverbials, which add information about the action or state described in the clause, answering questions such as 'How, When, Where, How much, To what extent?' and 'Why?' (ibid.: 763). The second type is the linking adverbials, which serve a connective function. They make the relationship between two units of discourse explicit. Lastly, the focus of this study is the third type: stance adverbials, classified by Biber into three subtypes: epistemic and attitudinal stance adverbials, and style stance. The latter is concerned mainly with conveying a speaker's comment on the style or form of the utterance, often clarifying how the speaker is speaking or how the utterance should be understood. However, epistemic stance adverbials convey certainty, doubt, and likelihood, whereas attitudinal stance adverbials convey a judgment or an evaluation (ibid.). As the study follows Biber and Finegan (1988), Biber et al. (1999) and Biber's (2006) framework on stance, expressions of stance adverbials are accordingly classified into epistemic and attitudinal markers. However, the only difference is that this research goes an extra mile, manually processing the semantic analysis of concordance lines to obtain meaningful results beyond frequency counts. To be able to clarify this part, two examples will be analysed. The two of them are extracted from the BNC2014 academic prose corpus. Both examples are of *perhaps*, which is classified as an epistemic adverbial of likelihood, showing that it fits under this classified category according to Biber et al. (1999):

3. is opposed to the 'rule' of bureaucracy, in spite and *perhaps* because of its unavoidable yet unintended promotion of bureaucratization Thus,

In the above example number 3, *perhaps* conveys a clear likelihood about the “rule of bureaucracy” without invoking any personal evaluation regarding that matter. *Perhaps* occurs as an epitemic marker in the vast majority of cases. However, it might occur as an attitudinal stance adverbial as in the following example:

4. half of which are distributed for free. In addition, a *perhaps* surprising number of people still pay for these, buying them

In 4, *perhaps* refers to a “number of people” that the writer evaluated/judged as “surprising”. Here, the classification of perhaps is treated as an attitudinal stance marker rather than an epistemic one, because it reflects the writer’s attitude.

3.2 Corpus Description

As explained above, the BNC1994 and BNC2014 represent a vital source for linguists who wish to study deep layers of language using the methodologies of corpus linguistics. In this study, (v 5.5.1) boxX was the only tool used to access both versions of BNC1994 and BNC2014. The extracted corpora from both BNC versions are: academic prose of 1994 and 2014 versions (henceforth: AP1994 and AP2014), and e-language register of 1994 and 2014 versions (henceforth: EL1994 and EL2014). Instances of stance adverbials that were ambiguous to classify as epitemic or attitudinal, most likely due to overlapping meanings or multifunctionality of the linguistic form, were labelled as no-stance. 2500 lines of concordance of AP1994, AP2014, EL2014, and 742 instances of EL1994 were analysed manually for semantic analysis, totalling 8242 lines of concordance throughout the study.

3.3 Data Extraction and Annotation

Primarily, all stance adverbs that were mentioned explicitly by Biber and Finegan (1988), Biber et al. (1999), and Biber (2006) were collected manually from these works. Then, they were examined one by one in the BNC registers via Lancsbox X software (v 5.5.1) to extract frequencies of stance across the four registers: academic prose registers AP1994 and AP2014 on the one hand, and e-language registers EL1994 and EL2014 on the other. Using LancsBox X (v 5.5.1) software, the concordance lines for each adverb were extracted and randomised in Python, selecting 500 random concordance lines of each examined adverb, marking 5 adverbs of each subcorpus, in total 20 adverbials across corpora, and finally saving the output to an Excel sheet to process each line manually for semantic analysis. Python software was once more used to perform the statistical tests and to produce the appropriate visualisations of the findings. The following tables show the 10 most frequent adverbials, normalised per 1M words (henceforth: pmw), as listed in Biber and Finegan (1988), Biber et al. (1999), and Biber (2006), sorted by lexical form and register. The full normalised and raw frequencies are available in Appendix 5.

Table 1 highlights the frequency distribution normalised pmw of epistemic and attitudinal adverbs explicitly mentioned by Biber and Finegan (1989), Biber et al. (1999) and Biber (2006) across the

academic prose and e-language registers of 1994 and 2014, respectively. Darker cell colour denotes higher frequency.

Table 1: the 10 most frequent epistemic and attitudinal adverbials (raw and normalised pmw) with darker colour indicating higher frequencies

EPISTEMIC ADVERBS									ATTITUDINAL ADVERBS								
Adverb	AP1994 n	AP1994 /1M	AP2014 n	AP2014 /1M	EL1994 n	EL1994 /1M	EL2014 n	EL2014 /1M	Adverb	AP1994 n	AP1994 /1M	AP2014 n	AP2014 /1M	EL1994 n	EL1994 /1M	EL2014 n	EL2014 /1M
As	134,080	8380.00	173,814	8690.70	1,309	6145.54	33,271	6654.20	Ironically	171	10.69	155	7.75	94	441.31	2,014	402.80
Like	9,476	592.25	8,804	440.20	489	2285.77	14,022	2804.40	Hopefully	84	5.25	110	5.50	41	192.49	727	145.40
Really	2,193	137.06	2,014	100.70	245	1150.23	10,689	2137.80	Unfortunately	653	40.81	499	24.95	26	122.07	675	135.00
Always	4,951	309.44	4,373	218.65	142	666.67	4,009	801.80	Importantly	326	20.38	1,224	61.20	7	32.86	155	31.00
Never	3,652	228.25	2,583	129.15	164	769.95	4,132	826.40	Sadly	103	6.44	53	2.65	11	51.64	382	76.40
Quite	3,891	243.19	2,436	121.80	166	779.34	3,027	605.40	Surprisingly	534	33.38	418	20.90	1	4.69	174	34.80
Probably	3,896	243.50	1,969	98.45	164	769.95	2,103	420.60	Rightly	293	18.31	193	9.65	4	18.78	57	11.40
Actually	2,557	159.81	2,415	120.75	87	408.45	2,808	561.60	Fortunately	151	9.44	92	4.60	2	9.39	97	19.40
Maybe	204	12.75	418	20.90	170	798.12	1,700	340.00	Even worse	46	2.88	44	2.20	4	18.78	47	9.40
Perhaps	5,403	337.69	3,990	199.50	77	361.50	893	178.60	Amazingly	11	0.69	8	0.40	2	9.39	85	17.00

The table demonstrates some interesting facts. The items appearing in the top 10 across all four registers in epistemic adverbs are *as*, *like*, *always*, *really* and *never*. None of these is a pure non-versatile, non-multifunctional, or non-ambiguous core epistemic stance adverbial, as they all exhibit high-frequency and multifunctional uses across registers. This means that no single epistemic stance adverbial dominates uniformly across registers and time periods: the epistemic adverbial system is register-differentiated rather than uniform, and the items that characterise epistemic stance in academic prose (*perhaps*, *indeed*, *clearly*, *probably*) are largely different from those that characterise spoken English (*really*, *maybe*, *definitely*, *actually*). Reading the normalised numbers above, as far as the corpora reflect, there is a huge gap between the frequency of epistemic and attitudinal adverbials: the former is distinctively higher in use across all registers.

The analysis of the samples was conducted manually using the Lancsbox X software (v 5.5.1). Academic prose AP1994, AP2014, and e-language EL2014 each contains 500 lines of concordance. The e-language register of BNC1994: EL1994 contains 774 lines of concordance only due to the register's small word count; only 213,000 words in total, with 8242 manually analysed concordance lines across four subcorpora. There have been ambiguous cases in which the stance adverbial could not be strictly classified; hence, it was labelled as a no-stance. The research adopted Biber's justification, where stance adverbials "can occur in a variety of positions in a clause". (Biber et al., 1999:770). In written discourse, the stance adverbial might enjoy an initial, medial, or final position

within a clause. This positional flexibility enables stance adverbials to function at the level of the entire clause (ibid). Adverbials with ambiguous scope are classified as (no-stance). This category includes stance expressions that do not meet the criteria for classification as either epistemic or attitudinal.

This thesis, as explained above, relies on qualitative and quantitative analyses to uncover a semantic layer and achieve more consistent results, rather than relying solely on frequency-based analysis. The following examples are analysed from the samples of the AP2014 register, where the adverb *indeed* shows the following various interpretations:

5. paper as an extended acknowledgement that Everettian quantum theory is indeed fundamentally flawed in its connection to experiment, since in their
6. row i, we also modify the sets to form sets indeed, is defined to consist of all vertices of, plus any
7. cells is wild-type, and not surprisingly, our study found that indeed, in 86% of cases, up-regulated p53 was wild-type, thus providing

In 5, the adverb *indeed* is evaluating the propositional clause “fundamentally flawed”; as such, it is classified as an attitudinal stance marker because it denotes an evaluation or judgment. However, in 6, the adverb *indeed* functions as an epistemic stance marker, as it expresses factual epistemicity. In 7, the scope of the adverb is ambiguous: it is not clear whether *indeed* refers to *our study indeed found*, or it refers to *86% of cases*. In such cases where semantic ambiguity is clear, the concordance line is classified as no-stance.

3.4 Data Analysis Procedure

The manually collected adverbs from Biber and Finegan (1988), Biber et al. (1999)I, and Biber (2006) were examined one by one on Lancsbox X (v 5.5.1) to obtain the frequency of each adverb. The list of every adverb was randomised in Python, selecting the first 500 concordance instances at random. Further, the output was saved into an Excel sheet file to manually process the data for semantic examination and classification into attitudinal or epistemic stance. The files are available in Appendices 1, 2, 3, and 4. Python was used to process the statistical tests and to produce the necessary data visualisations. Due to the difference in corpus size, normalisation is applied pmw. Registers are compared temporally, e.g. academic prose 2014 contrasts that of 1994, and so do the e-language registers in 2014 and 1994.

Using Python software, statistical tests were conducted on two types of analyses: Chi-square for statistical significance, and Cramér’s V for the effect size. The Chi-square test was used to correlate with specific registers and time periods whether the differences were statistically significant across the four subcorpora (academic prose and e-language in BNC1994 and BNC2014). This test is appropriate for categorical frequency data and was applied to determine the overall usage of stance categories between BNC1994 and BNC2014, as well as across academic prose and e-language. P-values indicate the significance of the Chi-square test results, with significance indicated by values

lower than 0.05. Cramér's V was calculated to evaluate the effect size. Collectively, these statistical tools provide a robust evaluation of stance adverbials over time and across registers.

3.5 Inclusion and Exclusion Criteria

The selection of stance adverbials in this thesis is a combination of theoretical grounding, empirical evidence, and cross-register comparability, rather than a purely frequency-based procedure. The candidate adverbs are first identified according to the functional categorisation proposed by Biber and Finegan (1988) and further developed by Biber et al. (1999) and Biber (2006). Together, the three works establish a very grounded approach to differentiate between epistemic and attitudinal expressions of stance. This ensures that all selected adverbs belong to a clearly defined grammatical approach and a clear semantic category; furthermore, it also ensures comparability across various registers and time periods. In a second step, information on corpus frequency was extracted from BNC1994 and BNC2014, as both corpora were used to identify the most frequent epistemic and attitudinal adverbs explicitly mentioned in the works of Biber's framework. This dual criterion allows the study to remain balanced in both empirical evidence and theoretical background.

Nonetheless, the study does not restrict the selection to the most frequent adverbs in each register or corpus. Such a restriction would show a methodological limitation within the study for two main reasons. First, frequency rankings are not indicative of the real use of epistemic and attitudinal stance adverbials, because they are not stable across registers or across time periods: an adverbial that ranks among the most frequent five in one register might fall outside this range in another, resulting in inconsistent datasets. Second, a ranking-based approach to adverbials would hinder the comparison of the same adverb across multiple registers and both time periods. Cross-register comparability is treated in this research as an essential selection criterion, as the central research objective is not the frequency count of each adverbial within each register in isolation, but to examine the different behaviour of the adverbs across registers and how they shift diachronically. As such, the study adopts a frequency-informed sampling approach, prioritising adverbials that occur across corpora and registers, rather than those that occupy a fixed rank within a single register or corpus. This selection measure assures that the selected adverbials are representative of recurring patterns of use and supports more reliable and systematic cross-corpus and cross-register comparison.

In addition to the criteria explained above, another important perspective is added: a functional and semantic one. Particular attention is given to adverbials that play a central role in the expression of stance and demonstrate either high functional stability or relevant analytical variability across registers. The epistemic category includes the following adverbials: *perhaps*, *indeed*, *clearly*, *probably*, *actually*, *really*, *maybe*, and *definitely*. They satisfy the multiple selection criteria simultaneously: they are explicitly mentioned and functionally salient in Biber's framework of stance

categorisation, and they are all frequent across registers representing a range of epistemic meanings. For instance, *probably*, *perhaps*, and *maybe* indicate varying degrees of probability and doubt or likelihood, whereas *definitely*, *indeed*, *really* and *actually* demonstrate certainty (Biber et al. 1999: 853,854; Biber (2006): 92). The adverb *clearly* is listed as a “surely adverb” that shows certainty (Biber and Finegan, 1988: 8). This categorisation of “surely adverbs” was further developed in Biber et al.’s (1999) framework to epistemic stance of certainty. Their collective inclusion guarantees that the analysis captures meaningful variation within the epistemic domain itself. These tokens also appear in sufficient varieties across both academic prose and e-language registers diachronically, which consequently enables cross-register comparison and the diachronic contrast that the study aims for. Some specific tokens, e.g., “*actually*” and “*probably*”, are used in the academic prose register. These two tokens appear frequently in the e-language register, and their inclusion in the academic prose was purposeful to enable comparability between the two registers of similar tokens rather than handling a completely different set for each register.

This research highlights the importance of semantic analysis. Throughout the manual analysis of the concordance lines, the semantic dimension was central to the process, yielding more authentic and consistent results. It is important, first, to understand why this semantic lens is crucial to the study. The answer comes directly after considering these three examples of *really* taken from the EL1994 corpus:

- 8. Nov 93 16:19:43 GMT Subject: Re: email footy Sorry this really is my last fantasy posting for a while, the R5
- 9. comfortable about losing the defensive midfielder. Point taken, and I really feel sorry for Rocky at the moment, and hope he
- 10. the new back pass rule last season. He did play really well in the Championship year but any time I saw

In 8, the adverb *really* indicates a fact. So, it is simply an epistemic stance marker of factuality. Whereas in 9, the speaker/writer conveys an attitude of being sorry about somebody, making *really* an attitudinal stance marker. Lastly, in 10, the speaker uses *really* as an intensifier of *well*, stressing how well a certain player played. As such, one adverbial *really*, while listed by Biber et al. (1999) and Biber (2006) as solely an epistemic stance marker of actuality, could play a multifunctional role in three different ways: epistemic stance, intensifier, and attitudinal stance. Douglas Biber himself, in his paper with Edward Finegan (1988), argues that

“After we had compiled a list of potential adverbial markers of stance, we set out to identify the occurrences of these adverbials that were actually functioning as stance markers in the texts. ... since most adverbial forms that function as stance markers can serve other functions as well” (1988: 8).

For this reason, and the previously detailed example findings, the study shows adverbials that demonstrate a higher degree of semantic flexibility, as this specificity is central to the research rather than incidental to it. That, in turn, means that the frequency analysis alone cannot capture this type of multifunctional stance-marking across registers and within the named time periods.

It is surprising that attitudinal stance marker counts are remarkably low in form and frequency across registers compared to epistemic ones. Because of their low number in form, there was no real need to divide the total number of analysed devices between epistemic and attitudinal stance markers. For this reason, “*unfortunately*” is the only retained attitudinal marker as a prototypical and empirically motivated candidate. The empirical motivation is evident from the frequency: it is the only attitudinal adverbial that maintains a consistent presence across all four registers and both time periods, ranking first in the AP1994 register, second in the AP2014 register, and third in both the EL1994 and the EL2014 registers. As provided in the frequency table in Appendix 5, no other attitudinal stance adverbial could achieve such comparable cross-register consistency, which makes “*unfortunately*” the best empirically evident choice among other candidates. Furthermore, “*unfortunately*” demonstrates a high degree of semantic stability that further strengthens its suitability as an analytical representative of the whole category. Unlike many epistemic adverbials that can shift function depending on context, “*unfortunately*” consistently expresses the speaker’s or writer’s negative evaluation of the propositional content without being ambiguous or overlapping with epistemic meaning. In the following example from Biber et al. (1999), *unfortunately* expresses, unambiguously, a clear attitudinal stance:

11. *Unfortunately*, I have too many of them, but someday you'll be old too (1999: 856).

In example 11, it is almost impossible to mistake the adverbial expression of the attitudinal stance. Such functional clarity, combined with its cross-register frequency, as highlighted earlier, makes the choice of *unfortunately* a well-grounded representative candidate of the attitudinal markers category. While providing an exhaustive account of all attitudinal adverbials in the corpora is not what this study aims to do, the inclusion of a prototypical, semantically stable, and frequent candidate such as “*unfortunately*” allows for a coherent comparison between the epistemic and attitudinal perspectives on stance within a uniform analytical framework.

Collectively, the study aims for a qualitative and quantitative approach to the analysis of epistemic and attitudinal stance adverbials, keeping both methodological and theoretical dimensions well-informed and principled. Thus, the selected adverbials exhibit sufficient frequency for statistical analysis and are semantically rich enough to support detailed contextual interpretation. The study spots the shift in stance use diachronically across registers by combining high-frequency epistemic markers that cover a range of epistemic meanings, along with richness in semantic meanings, with a prototypical attitudinal adverbial that demonstrates strong cross-register presence. This approach is based on the invaluable resources of Biber and Finegan (1988), Biber et al. (1999) and Biber (2006), where stance is not a fixed categorical lexical item, but a dynamic and context-relevant linguistic phenomenon shaped merely via the interaction between lexical choice, grammatical and syntactical

realisations, register, and function. A complete data sample with identified categories of stance (8,242 instances of adverbials) is attached to this thesis in the Appendices section.

3.6 Reliability and Validity measures

A corpus-driven approach was adopted to ensure the validity of the stance classifications used in the study. A total of 2500 concordance lines were manually examined per register, except the BNC e-language of 1994, where only 742 instances were analysed (the maximum number of concordance lines possible due to the register's low word count: only 213,000 words). The total number of concordance lines is 8242. Each of them was examined in context to determine its epistemic or attitudinal classification using manual semantic analysis. This process allowed for validating the stance categories across registers. Thus, the classification was based on actual usage rather than precompiled lists. Intra-coder agreement reached 90%.

4. Results and Discussion

This section represents the findings of the study regarding the analysis of stance adverbial distribution across the academic prose (AP) and e-language (EL) registers of the BNC 1994 and 2014. The analysis focuses on two categories: epistemic and attitudinal stance adverbials, whereas ambiguous instances were classified as no-stance. The following comparisons have been made: AP1994 vs AP2014, EL1994 vs EL2014. In addition to a full contrast between different registers comparison across the two corpora, AP1994 vs EL1994 and AP2014 vs EL2014. Due to the different sizes of corpora as mentioned in the previous sections, the frequencies of the stance adverbials were normalised to 1M words. This allows for a better comparison of frequencies across corpora of different sizes. The visualisations and the descriptive analysis both reflect the normalised frequencies. Statistical tests: Chi-square and Cramér's V are based on the raw count analyses as shown in Appendix 5.

4.1 Academic Prose: AP1994 vs AP2014 (Normalised pmw)

The normalised frequencies pmw in the Academic Prose (AP) register across the two corpora reveal a shift between BNC1994 and BNC2014. The AP1994 corpus demonstrates that the frequency of attitudinal adverbials is approximately 36.06, whereas the frequency of epistemic adverbials is 118.19. No-stance adverbials are significantly low at 2 pmw in the same corpus. On the other hand, the AP2014 corpus analysis shows a sharp decline in attitudinal frequencies to 8.7, epistemic adverbials remained almost stable at 112.2, while non-stance adverbials doubled, yet still at a low rate of 4.05 pmw.

Figure 1 illustrates this change in normalised frequencies pmw, demonstrating the sharp drop in attitudinal stance adverbials in the more recent corpus. The Chi-square test was performed on the raw counts, which showed that the difference in the distribution patterns between the two corpora is statistically significant (χ^2 (df=2, N=5000) = 267.804, $p < .001$), with a medium effect size ($V = 0.231$), exhibiting a diachronic change driven by the change in attitudinal adverbial usage rate.

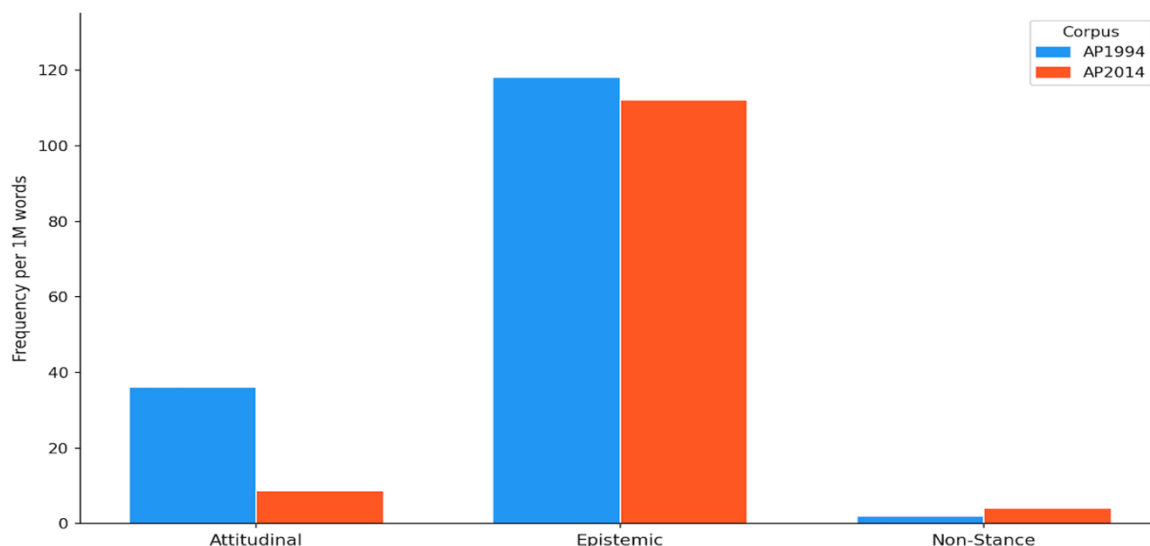


Figure 1: AP1994 vs AP2014 comparison (normalised pmw)

The figure shows a sharp drop in attitudinality in AP2014 against AP1994. Epistemic stance across the two corpora shows almost a stable performance. No-stance adverbials are very low. The shift is driven by, *unfortunately*, which exists in AP1994 and is lacking in AP2014.

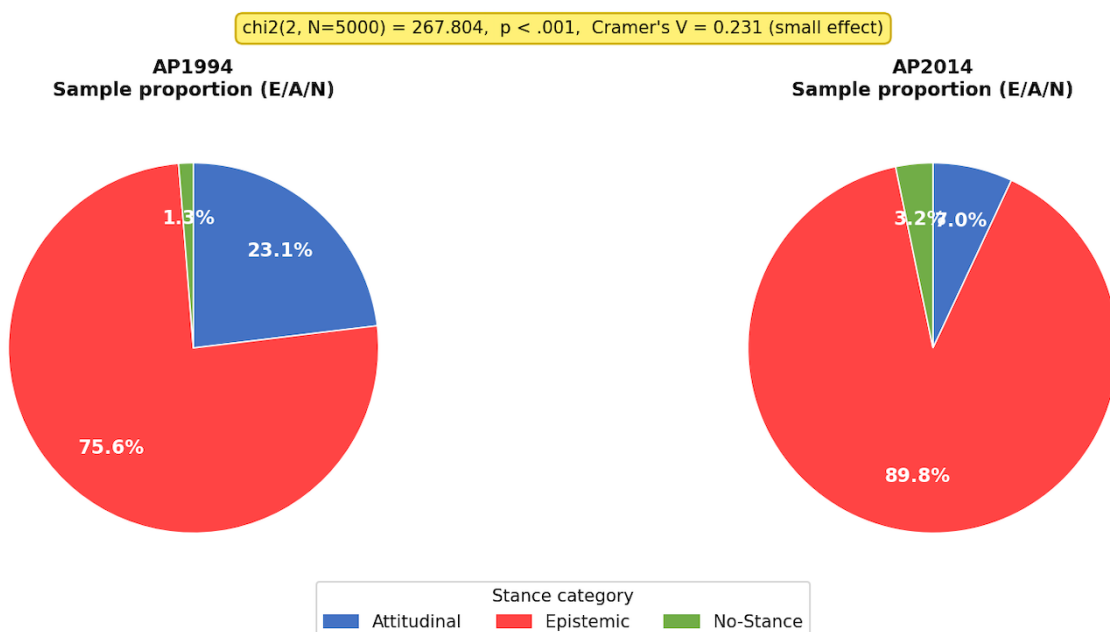


Figure 2: AP1994 vs AP2014 comparison in percentage

The percentage ratios in Figure 2 indicate that the epistemic stance remains almost the same, with a slight increase from 75.6% in AP1994 to 89.8% in AP2014. Whereas attitudinal stance drops to 7% in AP2014 compared to 23.1% in AP1994. The no-stance category remains very low in count.

It is noteworthy that the shift in AP2014 attitudinal adverbials is driven by the decline in the use of *indeed*, at 21.3, *perhaps* 20.75, and *clearly* 21.65 pmw each, compared to normalised 30.13, 29.38, and 29.5 in AP1994, respectively. In addition, the highest epistemic adverbs that show attitudinality in AP1994 are *probably*, *perhaps*, and *indeed*. Whereas AP2014 epistemic adverbs used attitudinally are *perhaps*, *indeed*, and *clearly*, in descending order, respectively, as shown in Appendix 5.

4.2 E-language: EL1994 vs EL2014 (Normalised pmw)

The normalised frequencies pmw in e-language show a dramatic difference and a significant diachronic change compared to the academic prose, particularly in the EL1994 sample. Due to the fact that the EL1994 register is a small corpus, as it counts only 213,000 words, attitudinal adverbs occurred at a very high rate of 474.18 pmw, and epistemic adverbials were frequent at 2,967.14. No-stance adverbials were also more frequent at 42.25. By contrast, the EL2014 corpus showed lower rates so far: attitudinal adverbials occurred at 104.2, epistemic adverbials at 387, and no-stance adverbials at 8.8 pmw.

Figure 3 illustrates these significant differences using normalised rates between the two e-language registers across the two corpora. Compared with the observed rates, usage of all stance types declines, especially epistemic. The Chi-square test on raw counts indicated a statistically significant difference between the two e-language samples (χ^2 (df=2, N=3,242 = 21.00, $p < .001$); the effect size based on raw counts was small ($V = 0.081$). The normalised frequencies reveal a substantial change in the frequency, likely influenced by the nature and size of the EL1994 sample.

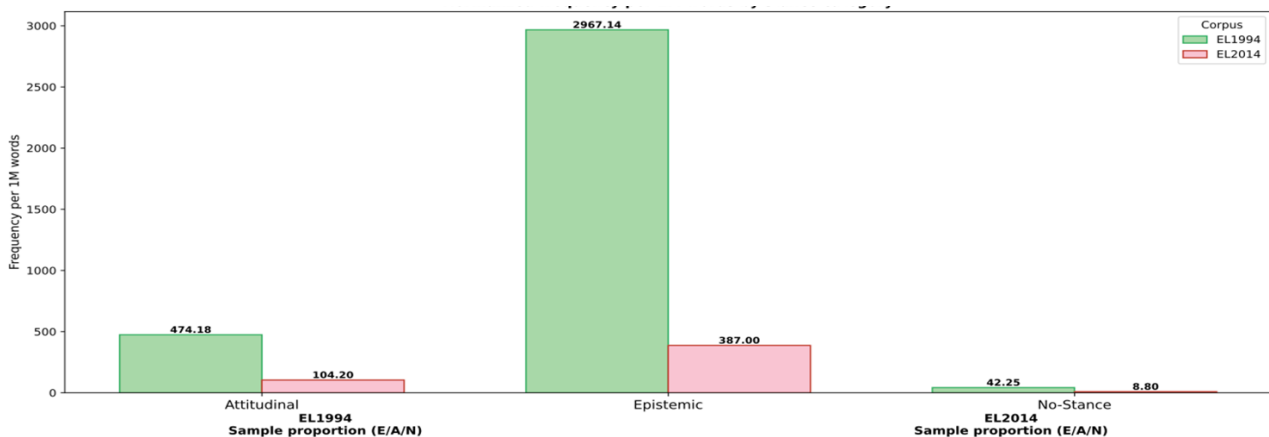


Figure 3: EL1994 vs EL2014 comparison (normalised pmw)

The figure shows that attitudinal and epistemic adverbs of EL1994 are much higher in frequency than their counterparts in EL2014. The small size of the EL1994 corpus might be the reason for such a result. The most frequent epistemic adverbs, normalised pmw, in EL2014 are *maybe* at 93.6, *actually* at 86.2 and *probably* at 84.8, respectively. Epistemic adverbs that show attitudinal use are *really*, *definitely*, and *probably*, as demonstrated in Appendix 5.

The results show the diachronic change in the use of epistemic and attitudinal stance adverbials between EL1994 and EL2014. Due to the small sample size of EL1994, the results show inconsistency. However, the most frequent adverbial use of epistemic adverbs in EL2014 are *maybe* at 93.6 per 1M words, *actually* at 86.2, and *probably* at 84.8. The most epistemic adverbs that were used attitudinally are *really*, *definitely* and *probably* in descending order, respectively, as shown in Appendix 5.

4.3 Cross-register comparison across corpora

4.3.1 AP1994 vs EL1994 (Normalised pmw)

In BNC1994, the EL1994 corpus, previously highlighted for its low word count of 213,000, showed higher normalised frequencies pmw across all stance types than in AP1994. Attitudinal adverbials were used almost 13 times more frequently in e-language (474.18 vs 36.06). Epistemic adverbials were 25 times more frequent (2967.14 vs 118.19). Non-stance adverbials occurred approximately 21 times more frequently (42.25 vs 2). Chi-square test shows a significant difference in distribution (χ^2 , $df=2$, $N=3,242$) = 31.23, $p < .001$); however, a small effect size was noted ($V=0.098$).

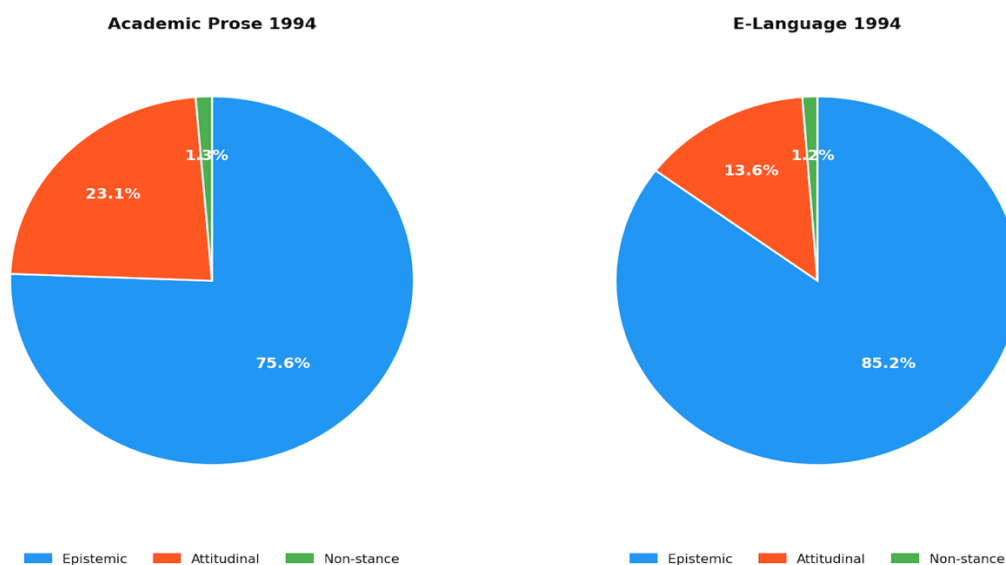


Figure 4: cross-register comparison of AP1994 vs EL1994 (normalised pmw)

Figure 4 shows that EL1994 has higher absolute frequencies across all categories (as noted above). Epistemic adverbials constitute a larger proportion of its reported stance type (85.2% vs 75.6%), while attitudinal adverbials represent a greater share in AP1994 (23.1% vs 13.6%), reflecting different functional priorities and user preferences across the two sub-corpora.

4.3.2 AP2014 vs EL2014 (Normalised pmw)

In BNC2014, the EL2014 register exhibited higher normalised frequencies than AP2014, but the gap was significantly smaller than in BNC1994, likely due to the smaller sample size. Attitudinal adverbials in normalised frequency pmw were used approximately 10 times more frequently in the EL2014 register (104.2 vs 8.7). Epistemic adverbials were almost 3.5 times more frequent (387 vs 112.2), whereas non-stance adverbials almost doubled in frequency (8.8 vs 4.05). Chi-square test on raw values indicated a highly significant difference in distribution (χ^2 , $df=2$, $N=5000$) = 207.193, $p < .001$), with a small effect size ($V = 0.204$), indicating a distinct difference in usage patterns.

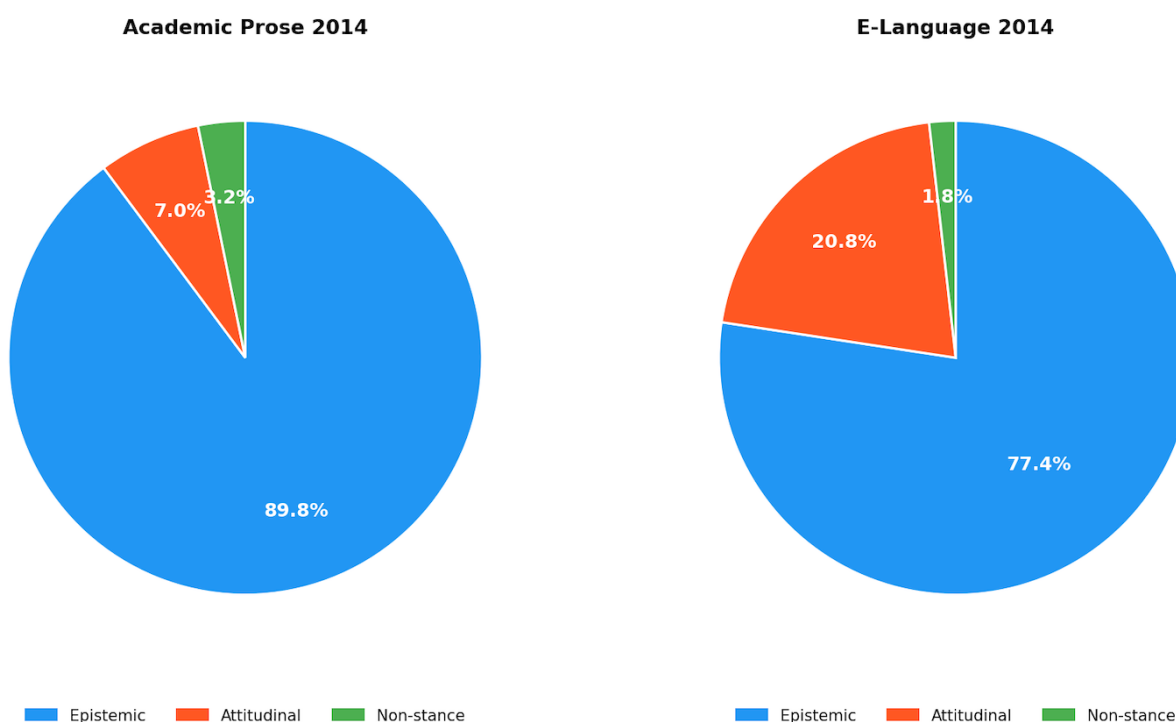


Figure 5: a cross-register comparison of AP2014 vs EL2014 (normalised pmw)

The figure demonstrates that while epistemic stance dominates both registers (89.8% and 77.4%, respectively), E-Language exhibits a notably higher attitudinal stance (20.8%) than Academic Prose (7.0%), suggesting more personal evaluation and affect in electronic communication. Non-stance markers remain minimal in both contexts.

4.3.3 Overall Normalised Distribution Across Corpora

An overall analysis encompassing all four registers confirms significant variation in distribution patterns based on raw values (χ^2 , $df = 6$, $N = 8242$) = 300.030, $p < .001$), with a small-to-medium overall effect size ($V = 0.135$). Normalised frequencies pmw provide a clearer image of the differences in rates.

Figure 6 represents a grouped bar chart contrasting the normalised frequencies pmw for each stance type across all four sub-corpora. This visualisation highlights significantly high rates in the EL1994 sample, for both types of adverbials. It shows lower rates of academic prose than e-language over the two decades, with a decrease in attitudinal adverbials in both the EL2014 and AP2014 subcorpora.

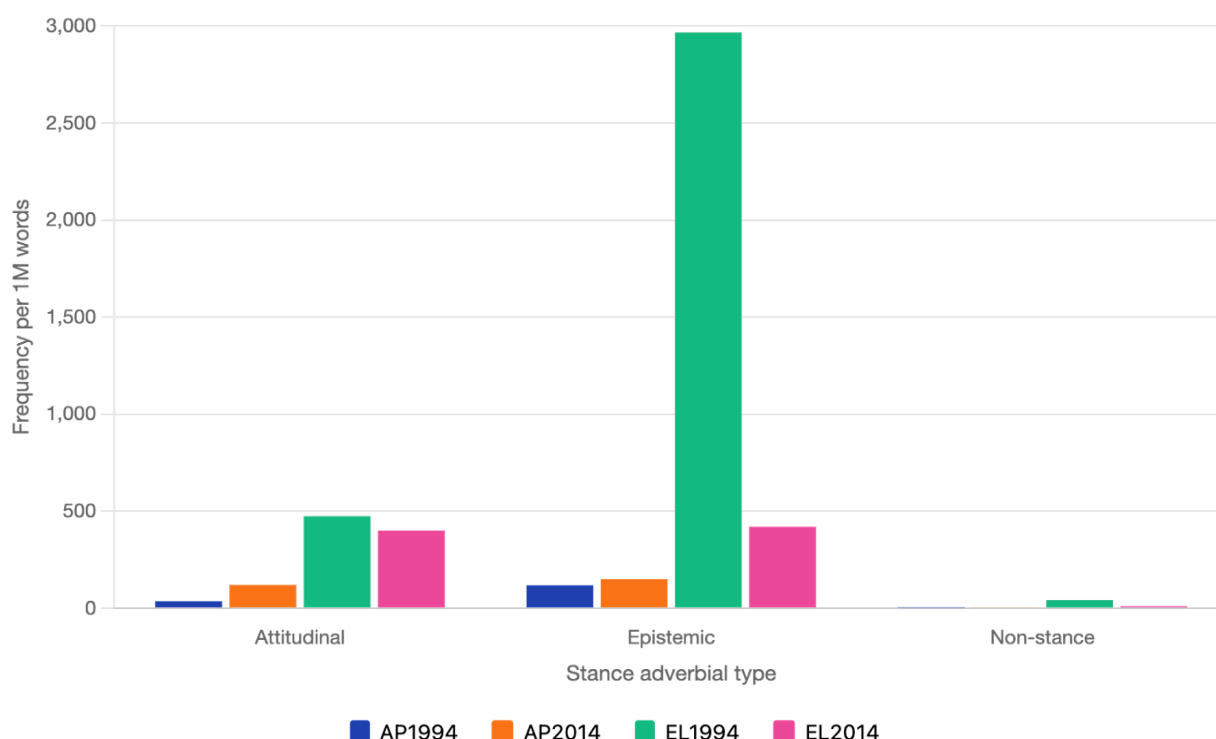


Figure 6: a bar chart showing an overall normalised distribution across corpora

The figure demonstrates stance adverbial frequency across all corpora normalised pmw. The low word count of EL1994 shows the highest frequency across all corpora, particularly for epistemic stance (2,967 pmw), while both Academic Prose corpora show remarkably lower frequencies. No-stance adverbials remain minimal across all corpora.

Figure 7 shows a heat map that provides a clear view of the normalised frequencies, with colour intensity representing the rate pmw. The high frequency of epistemic adverbials in EL1994 is clear

in intense pink colour. It is also clear how the non-stance shows a very low frequency across all registers. Meanwhile, attitudinal adverbials occur infrequently in AP2014.

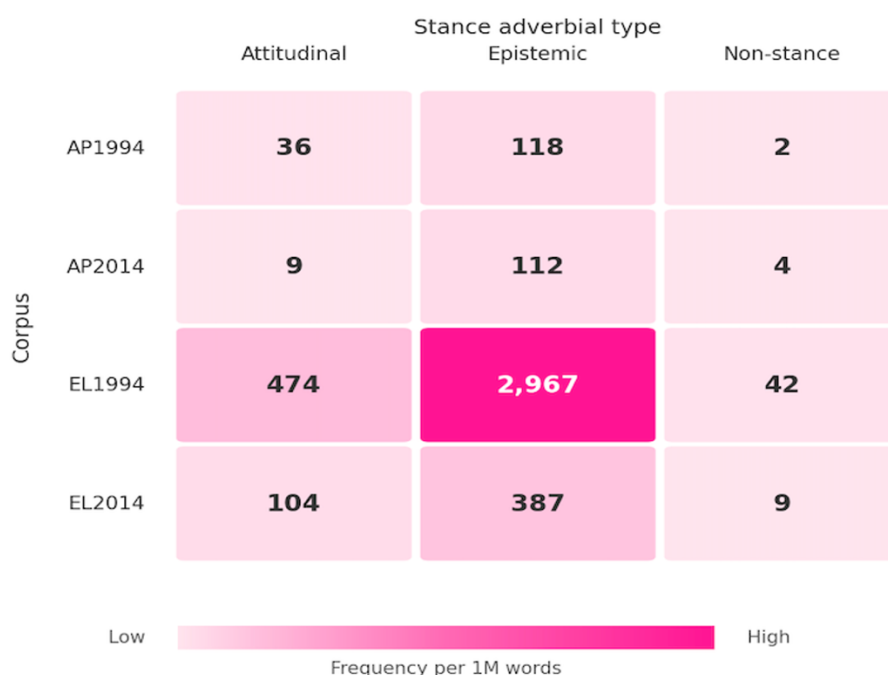


Figure 7: a heatmap showing an overall normalised distribution across corpora

The figure demonstrates a heatmap of normalised stance adverbial frequency across corpora and stance types. Distribution of stance adverbial frequencies, normalised pmw, across the four corpora (AP1994, AP2014, EL1994, EL2014) and three stance adverbial categories (Attitudinal, Epistemic, No-stance). Colour intensity indicates frequency, with darker pink representing higher rates of occurrence.

4.4 Discussion

This MA thesis investigated the distribution of stance adverbials across academic prose and e-language registers in the BNC1994 and BNC2014 corpora, with four research questions guiding the analysis. The findings reveal interesting patterns that highlight synchronic register variation and diachronic linguistic change: two different patterns that can be interpreted in the light of Halliday's register model and Traugott's subjectification theory, respectively.

The first research question examined whether the use of stance adverbials across matched registers differs across time periods in the BNC1994 and BNC2014. The data show that yes, there are some differences. In academic prose, the change involves a distinct decline in the attitudinal adverbials from 36 to 8.7 pmw: a 78% decrease in normalised frequency. This change is statistically

significant, with a medium effect size ($df = 2$, $\chi^2 = 267.80$, $p < .001$, $V = 0.231$). This change demonstrates how academic writers use stance markers in their texts. In contrast, epistemic adverbials remained stable at 118 vs 112.25 pmw, suggesting that certainty markers continue to serve essential writing functions in the academic written discourse.

Across the e-language registers, all stance types were less frequent in the EL2014 than in the EL1994. The EL1994 sample exhibited extremely high frequency rates, which may be a limitation given its low word count (only 213,000 words). The EL2014 shows more balanced data: with epistemic adverbials at 387 and attitudinal at 104.2 pmw. The EL1994 shows higher rates of stance markers than the EL2014, suggesting differences in how interpersonal meanings are expressed across contexts. The diachronic comparison reveals convergence between registers.

The second research question concerns whether diachronic change manifests differently across registers. The findings confirm a change in this aspect as well. Academic prose registers across time periods show a decline in the attitudinal stance-marking against the epistemic markers, suggesting that evaluative subjectivity is less favoured. This tendency reflects the register's particular priorities. As Halliday and Matthiessen (2014) argue, register variation emerges from "systematic correspondences between the contextual values and the meanings that are at risk in the contexts defined by these values" (2014:34). In academic prose, the tenor is institutionally distant, and the mode favours the ideational meaning over the interpersonal positioning. As such, attitudinality conflicts with the objectivity required by academicians. However, epistemic markers align with the academic requirements of certainty, evidentiality, and claims of knowledge. E-language registers display a contrasting pattern. Although the EL1994 data may not provide a reliable interpretation due to limited word count, the EL2014 data indicate that attitudinality remains central. The tenor of e-language is informal and high-contact, with writers addressing readers directly and personally, often due to established familiarity. In this discursive context, stance is essential, as Halliday and Matthiessen emphasise: "the combinations of field, tenor and mode values determine different uses of language" (2014:34).

This means that stance, in such a context, becomes highly required because of the combinations of field, tenor and mode as explained, denoting that subjective positioning becomes indispensable in such a context. Additionally, the statistical data reveal that a higher frequency of all stance types in EL2014 compared to AP2014 epistemic: 387 vs 112.2; attitudinal: 104.2 vs 8.7, confirms that this register favours interpersonal, subjective meaning markers rather than its academic prose counterpart.

The third research question concerns the interpretation of the findings in light of theories and models of language evolution. The study's findings can be interpreted with Traugott's theory of subjectification, which is defined as "Meanings tend to become increasingly based in the speaker's subjective belief state/attitude toward the proposition" (Traugott, 1989:35). Traugott points out that

there is a gradual change in which overt, objective meanings transform into subjective, interpersonal ones over time. This phenomenon aligns with the cross-register comparison between academic prose and e-language registers. EL registers show a higher frequency of stance than AP across both corpora. In EL1994, the attitudinal and epistemic expressions of stance exhibit a normalised frequency of 474 and 2,967.14, respectively, compared to 36 and 118, respectively, in AP1994. Furthermore, in EL2014, attitudinal and epistemic markers demonstrate a shift at 104.2 and 387, respectively, compared to 104.2 and 8.7 and 112.25 in AP2014, respectively. The pattern reflects the EL's tendency toward more subjective interpersonal discourse. Whereas AP is governed by the objectivity conventions and academic norms that prioritise objective distant discourse.

By contrast, a counter-subjectification phenomenon can be observed in the academic prose register diachronically. Attitudinal markers declined from 36.1 to 8.7 pmw between 1994 and 2014, representing a 78% decrease in normalised frequency. This is statistically significant, with a meaningful effect size ($df = 2$, $\chi^2 = 267.804$, $p < .001$, $V = 0.231$). Academic writers are not shifting stance to other linguistic constructions, but they are reducing the overt marking of interpersonal, evaluative self-positioning. This pattern of academic prose shows that academic writing is driven by institutionality and other factors that govern the shift.

As such, that divergence in patterns across registers is diachronic: a comparison between EL and AP registers aligns with Traugott's (1989) subjectification theory, whereas academic prose resists it, thereby supporting a better understanding of language change. Subjectification, in this context, captures important trends in cross-register comparison; whereas specialised registers such as academic prose are governed by institutional and academic standards driven by the pressure of supervisors and academic conventions. Therefore, registers are not only a language-mirroring device that exhibits synchronic changes as a static entity, but they are also dynamic, continuously changing, shaping the direction of diachronic change as well.

There is a pattern of *conversion* between epistemicity and attitudinality that can be observed through the analysis of epistemic and attitudinal stance adverbials. It was noted that it is easy and frequent for epistemic adverbials to refer to attitudinality, while vice versa is very difficult. From the analysed concordance lines, it was observed that if an evaluative content word appears in the proposition after the adverb, the meaning is likely to convert toward attitudinality. Perhaps this *convertibility* is the reason that makes speakers and writers prefer epistemic adverbials and convert them to attitudinal, given the limitedness of both form and function of attitudinal adverbial markers, as shown in Appendix 5, compared to the large number of linguistic devices used to convey epistemic meanings. As such, this would give language users greater freedom to use a linguistic device that can be easily *converted* rather than a static linguistic form, which cannot be easily converted, such as the attitudinal adverbials.

5. Conclusion

This MA thesis investigates the diachronic change of stance markers, particularly epistemic and attitudinal stance adverbials, across two registers: academic prose and e-language in the BNC1994 and the BNC2014. By manually analysing 8242 lines of concordance lines and normalising the frequency across corpora of varying sizes pmw, the study could provide strong evidence of how stance markers in British English evolved over time.

The findings reveal a very clear register-specific trend: while the e-language register regards epistemic and attitudinal stance adverbials as essential for a complete communication aligning with its register-specific requirement, specifically its dialogic and interpersonally rich tenor, academic prose marks a drop in attitudinal stance expressions for epistemic ones, likely increasing the register-specific constraints toward a more overtly objective and impersonal academic writing tone. These findings support the theoretical framework of the register model in Halliday's systemic functional grammar (2014), demonstrating that contextual factors help shape the linguistic realisation of stance expressions. In addition, the findings support Traugott's (1989) subjectification theory, highlighting that a cross-register comparison is consistent with the framework of subjectification, in contrast to a counter-subjectification trend in diachronic academic prose registers that shows resistance to Traugott's (1989) framework.

Not only does this thesis represent a descriptive contribution, but it also has other implications for multiple audiences. For scholars in corpus linguistics and discourse analysis, it offers insights into tracing linguistic variation through normalised frequencies, semantic analysis, and register-oriented comparisons. For students and teachers of English as a Second Language (ESL), the study offers useful insights into the need to avoid personal and subjective stance expressions, especially in academic writing. Finally, the study contributes to the public understanding of language change and evolution, showing how, in particular, the genres of digital communication and academic writing evolve, specifically in how speakers and writers express themselves through language.

While the study provides valuable insights into language change, it is not without limitations. The rigid selection criteria for adverbs, e.g., explicitly mentioned in the framework, could be less restrictive and include other adverbials that could yield other insightful results. In addition, the sparsity of the e-language corpus of BNC1994 affected the register comparisons, even with normalised values. Moreover, the manual analysis of 8242 concordance instances, despite being methodologically grounded, might introduce judgmental interpretation, leading to subjectivity. The study was limited to five adverbials from each subcorpus, extracted relatively in a large number of instances, thus increasing the number of adverbials or the number of the extracted instances might help show diverse outputs. Examining many epistemic adverbials was enough to observe the patterns, whereas attitudinal adverbials were not sufficiently examined through the analysis. Lastly, the

analysis investigated only stance adverbials; however, other grammatical markers, such as modal verbs and style stance, could be examined by other researchers.

Future investigations have significant potential for further scholarly exploration in the same study scope. Depending on a different framework would yield different results. Additionally, conducting cross-linguistic studies between English and other languages might help determine whether these diachronic patterns are generalisable to other languages. As the linguistic scope continues to expand, especially in digital communication, the way writers and speakers position themselves in speech or writing will remain a competitive area of research.

Summary

This MA thesis is a corpus-based study of stance adverbials in British English, using the BNC1994 and BNC2014 corpora to investigate diachronic changes, specifically regarding epistemic and attitudinal stance adverbials. The study is grounded in Biber and Finegan (1989), Biber et al. (1999) and Biber's (2006) framework of stance. It is focused on the academic prose and e-language registers of the two periods. Totally, there are four sub-corpora examined: Academic Prose 1994 and 2014 (henceforth AP1994 and AP2014), and e-language 1994 and 2014 (henceforth EL1994 and EL2014).

After scanning the books of Biber's framework, all the adverbials were collected, according to his classification into epistemic and attitudinal adverbs. Then, using Lancsbox X (v 5.5.1), each adverb was scanned for frequency. In the next step, the concordance lines for each adverb were processed in Python to randomise, selecting the first 500 lines to analyse manually. The analysis in this study goes beyond frequency counts and analyses, as it extends to the semantic analysis necessary to determine the adverbial's true category: epistemic or attitudinal. The study applies two statistical tests to the findings: Chi-square tests to assess significance and Creamér V to estimate effect size. As such, this research operationalises both quantitative and qualitative methods. The explored patterns are anchored to Traugott's (1989) theory of subjectification and Halliday's (2014) register theory.

The analysis of stance adverbials across the BNC1994 and BNC2014 corpora reveals significant diachronic shifts and register-specific patterns. In academic prose, attitudinal adverbials decreased sharply from 36.06 to 9.95 per 1M words between 1994 and 2014. While epistemic adverbials remained relatively stable at approximately 111-118 per 1M words. E-language exhibited higher frequencies in both periods (epistemic: 2,967.14 per 1M words; attitudinal: 474.18 per 1M): likely due to the low word count of the EL1994 corpus compared to more moderate frequencies in the EL2014 (epistemic: 387; attitudinal: 104.2). Attitudinal stance was more pervasive in the e-language register, whereas academic prose registers showed high frequency of epistemicity across corpora. There are ambiguous cases, which could not be classified as epistemic or attitudinal: for these, a no-stance label was added

This thesis represents a new layer in the manual analysis of stance: the semantic dimension. That shows the importance of this deep analysis step for more consistent results. Generally speaking, epistemic adverbials show a higher rate of convertibility into attitudinal ones. A speaker/writer can add content words, e.g. adverbs, nouns or verbs that hold an evaluation or assessment to the proposition preceded or followed by an epistemic stance adverbial, nuancing the meaning of the epistemic adverbial and converting it to an attitudinal expression. However, through the analysis, the conversion of attitudinal into epistemic is rare. However, some limitations can be spotted. The small

word count of the EL1994 register indicated inconsistent results. Additionally, the manual analysis is subjective and vulnerable to error. Despite the fact that the analysed lines are 8242 concordance lines, they still resemble a small sample, given the high number of adverbs across corpora. This study is a basis for future researchers who would like to investigate further the semantic dimension of stance adverbials and the convertibility exhibited by epistemic adverbs across corpora. There are implications for ESL teachers, who can use the findings to better guide MA and PhD students in observing and writing clearer English that aligns with the academic conventions of overt objectivity and less subjectivity. Additionally, the study provides useful insights into the general understanding of language change diachronically.

Santrauka

Šis magistro darbas yra tekstyninis autoriaus poziciją reiškiančių adverbialų tyrimas britų anglų kalboje, naudojant BNC1994 ir BNC2014 tekstynus ir siekiant ištirti diachronines permainas, ypač susijusias su epistemiškumo ir požiūrio raiška. Tyrimas grindžiamas Biber ir Finegan (1989), Biber *et al.* (1999) bei Biber (2006) teorine prieiga. Jis orientuotas į akademinės anglų kalbos ir elektroninės kalbos registrus iš dviejų laikotarpių. Iš viso tiriama keturi patekstyniai: akademinė rašytinė kalba iš abiejų tekstynų (toliau tekste – AP1994 ir AP2014) bei elektroninė kalba iš abiejų tekstynų (EL1994 ir EL2014). Išnagrinėjus Biber ir kitų tyrėjų mokslinius darbus, buvo sudarytas tyrimui reikalingas adverbialų sąrašas. Naudojant programą „Lancsbox X“, sukaupiti duomenys apie kiekvieno adverbialo dažnumą kiekviename patekstynyje. Kitame etape kiekvieno adverbialo konkordansų eilutės buvo apdorotos „Python“ programa ir sudarytos atsitiktinės 500 eilučių imtys kokybinei analizei. Kiekybinę analizę papildė semantinė analizė, siekiant nustatyti adverbialų kategorijas: episteminę ar požiūrio. Tyrime taikyti du statistiniai testai: Chi-kvadrato testas reikšmingumui įvertinti ir Cramerio V testas efekto dydžiui apskaičiuoti. Tyrimo rezultatai aptariami Traugott (1989) subjektyvėjimo teorija ir Halliday (2014) modalumo teorijų kontekste.

Adverbialų analizė BNC1994 ir BNC2014 tekstynuose atskleidžia reikšmingus diachroninius pokyčius ir registrams būdingus dėsningumus. Akademinėje kalboje požiūrio raiškos adverbialų per tirtą laikotarpį sumažėjo nuo 36,06 iki 9,95 viename milijonui žodžių, tuo tarpu epistemiškumo raiškos adverbialai išliko santykinai stabilūs (apie 111-118 atvejų viename milijonui žodžių). Elektroninė kalba parodė aukštesnius dažnius abiem laikotarpiais (episteminiai: 2 967,14 viename milijonui žodžių; požiūrio: 474,18 viename milijonui žodžių). Tikėtina, dėl mažo EL1994 korpuso žodžių skaičiaus, elektroniniame registre dažniai žemesni (episteminiai: 387; požiūrio: 104,2). Požiūrio raiškos adverbialai dažnesni elektroninės kalbos registre, tuo tarpu akademinės prozos registrai parodė didelį epistemiškumo adverbialų dažnį. Tyrime nustatyta ir dviprasmiškų atvejų, kurių negalima klasifikuoti pagal pasirinktas kategorijas; tokie advejai klasifikuoti kaip „Nežymėti“.

Šis darbas apima ir semantinę analizę. Tyrimas patvirtina, kad episteminiai adverbialai gali būti dažniau vartojami reikšti autoriaus požiūriui. Tuo tarpu požiūrio raiškos adverbialai retai vartojami epistemiškumui reikšti.

Šis tyrimas turi tam tikrų ribotumų. Mažas EL1994 patekstynio žodžių skaičius neleidžia gauti nuoseklius rezultatus. Be to, rankinė kokybinė analizė išlieka gana subjektyvi. Nepaisant nemažos tyrimo imties (8242 konkordanso eilutės), šis darbas vis tiek išlieka ribotas. Todėl ateityje jis galėtų būti plečiamas, apimant daugiau skirtingų patekstynų ir kalbos atmainų. Autoriaus pozicijos raiškos tyrimai gali praversti svetimkalbių studentų kalbinės kompetencijos ugdymui, siekiant geriau parengti juos magistrantūros ir doktorantūros studijoms. Be to, tyrimas suteikia naudingų įžvalgų į diachroninę kalbos raidą.

References

Primary References:

BNC website (no date) 'Genres'. Available at: https://gawron.sdsu.edu/functions_of_language/course_core/lectures/genres.html (Accessed: 2 May 2026).
British National Corpus 2014 (no date) Available at: <http://corpora.lancs.ac.uk/bnc2014> (Accessed: 30 January 2026).

Software:

Lancbox X (v. 5.5.1)

Python

Secondary References:

Abdeljaoued, M. (2026) 'From hedging to self-mention: Seventy years of stance in doctoral theses and what it means for EAP', *Journal of English for Academic Purposes*, 79, 101616. Doi: 10.1016/j.jeap.2025.101616.

Allan, K. (2013) *The Oxford handbook of the history of linguistics*. Oxford: Oxford University Press.

Ağçam, R. (2015) 'A corpus-based study on attitudinal stance in native and non-native academic writing', *International Journal of Humanities Social Sciences and Education (IJHSSE)*, 2(8), pp. 123-129.

Banaitytė, A. (2017) *Author stance in argumentative journalistic discourse*. BA thesis. Vilnius University.

Aijmer, K., 2020. That's absolutely fine: an investigation of absolutely in the spoken BNC2014. In: V. Brezina, R. Love and K. Aijmer, eds. *Corpus approaches to contemporary British speech: sociolinguistic studies of the spoken BNC2014*. London: Routledge, pp.143–167

Biber, D. et al. (1999) *Longman grammar of spoken and written English*. Harlow: Pearson Education Limited.

Biber, D. (2006) *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

Douglas Biber & Edward Finegan (1988) Adverbial stance types in English, *Discourse Processes*, 11:1, 1-34, DOI: 10.1080/01638538809544689

Brezina, V., McEnery, T. and Wattam, S. (2020) *LancsBox X: A software tool for corpus linguistics (Version 5.5.1) [Computer software]*. Lancaster: Lancaster University. Available at: <https://corpora.lancs.ac.uk/lancsbox/> (Accessed: 2 May 2026).

- Dugalich, N.M. and Han, H. (2024) 'Certainty stance adverbs in Chinese linguistic academic writing: A corpus-based study', *RUDN Journal of Language Studies, Semiotics and Semantics*, 15(1), pp. 248-261. Doi: 10.22363/2313-2299-2024-15-1-248-261.
- Halliday, M.A.K. and Matthiessen, C.M.I.M. (2014) *Halliday's Introduction to Functional Grammar*. 4th edn. Abingdon: Routledge.
- Hunston, S. (2011) *Corpus approaches to evaluation: Phraseology and evaluative language*. New York: Routledge, Taylor & Francis Group.
- Hunston, S. (2022) *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. and Thompson, G. (eds.) (2000) *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press
- Hunston, S. and Thompson, G. (eds.) (2000) *Evaluation in text: authorial stance and the construction of discourse*. Oxford: Oxford University Press.
- Hyland, K. (2018) *The essential Hyland: Studies in applied linguistics*. London: Bloomsbury Academic.
- Hyland, K. (2019) *Metadiscourse: exploring interaction in writing*. 2nd edn. London: Bloomsbury Academic.
- Kaatari, H. and Larsson, T., 2019. Using the BNC and the spoken BNC2014 to study the syntactic development of I think and I'm sure. *English Studies*, 100(6), pp.710–727
- Kuhi, D. and Rezaei, S. (2020) 'Diachronic Analysis of Stance Markers in Research Articles' discussion Sections', *Khazar Journal of Humanities and Social Sciences*, 23(4), pp. 33-46. Doi: 10.5782/2223-2621.2020.23.4.33.
- Laws, J., Ryder, C. and Jaworska, S., 2017. A diachronic corpus-based study into the effects of age and gender on the usage patterns of verb-forming suffixation in spoken British English. *International Journal of Corpus Linguistics*, 22(3), pp.375–402.
- Leech, G.N., Rayson, P. and Wilson, A. (2001) *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Routledge.
- Martin, J.R. and White, P.R.R. (2005) *The language of evaluation: Appraisal in English*. Basingstoke: Palgrave Macmillan.
- Meng, Q.L., 2019. A corpus-based investigation of diachronic evolution of like and its variants through grammaticalization. *Advances in Language and Literary Studies*, 10(3), pp.81–88
- Poole, R., Gnann, A. and Hahn-Powell, G. (2019) 'Epistemic stance and the construction of knowledge in science writing: A diachronic corpus study', *Journal of English for Academic Purposes*, 42, 100784. Doi: 10.1016/j.jeap.2019.100784.
- Poole, R., Gnann, A. and Hahn-Powell, G. (2021) 'Corrigendum to "Epistemic stance and the construction of knowledge in science writing: A diachronic corpus study."' [Journal of English

- for Academic Purposes (2019), Volume 42, 100784]’, *Journal of English for Academic Purposes*, 50, 100959. Doi: 10.1016/j.jeap.2021.100959.
- Simon-Vandenberg, A.-M. (2008) ‘Almost certainly and most definitely: Degree modifiers and epistemic stance’, *Journal of Pragmatics*, 40(8), pp. 1521-1542. Doi: 10.1016/j.pragma.2008.04.015.
- Sinclair, J. (1991) *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Tantucci, V. (2015) ‘Epistemic inclination and factualization: a synchronic and diachronic study on the semantic gradience of factuality’, *Language and Cognition*, 7(3), pp. 371-414. Doi: 10.1017/langcog.2014.34.
- Traugott, E.C. (1989) ‘On the rise of epistemic meanings in English: An example of subjectification in semantic change’, *Language*, 65(1), pp. 31-55. Available at: <http://www.jstor.org/stable/414841> (Accessed: 15 July 2014).

Appendices

Appendix 1: Academic Prose 1994 data.

Appendix 2: Academic Prose 2014 data.

Appendix 3: E-language 1994 data.

Appendix 4: E-language 2014 data.

Appendix 5: Excel sheet of the raw and normalised pmw frequencies, statistical tests for selected adverbials, and all the reference-listed adverbials' raw and normalised pmw frequencies.

Appendix 5

Selected and Analysed Stance Adverbials: Combined Corpus Comparison (1994 vs 2014)

A = Attitudinal | E = Epistemic | N = Neutral | — = Not examined

Word	AP 1994			AP 2014			EL 1994			EL 2014		
	A	E	N	A	E	N	A	E	N	A	E	N
Actually	—	—	—	3	482	15	8	79	0	64	431	5
Clearly	6	472	22	18	433	49	—	—	—	—	—	—
Definitely	—	—	—	—	—	—	—	—	—	121	375	4
Indeed	18	482	0	71	426	3	—	—	—	—	—	—
Maybe	—	—	—	—	—	—	8	158	4	16	468	16
Perhaps	25	470	5	77	415	8	7	69	1	—	—	—
Probably	28	467	5	5	489	6	23	140	0	69	424	7
Really	—	—	—	—	—	—	55	186	4	251	237	12
Unfortunately	500	0	0	—	—	—	—	—	—	—	—	—
TOTAL (raw)	577	1891	32	174	2245	81	101	632	9	521	1935	45
Per 1M words	36.1	118.2	2	8.7	112.2	4	474.2	2967.1	42.3	104.2	387	9

Appendix 5

[illegible]

AP 1994 /1M			AP 2014 /1M			EL 1994 /1M			EL 2014 /1M		
A	E	N	A	E	N	A	E	N	A	E	N
—	—	—	0.15	24.10	0.75	37.56	370.89	0.00	12.80	86.20	1.00
0.38	29.50	1.38	0.90	21.65	2.45	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	24.20	75.00	0.80
1.13	30.13	0.00	3.55	21.30	0.15	—	—	—	—	—	—
—	—	—	—	—	—	37.56	741.78	18.78	3.20	93.60	3.20
1.56	29.38	0.31	3.85	20.75	0.40	32.86	323.94	4.69	—	—	—
1.75	29.19	0.31	0.25	24.45	0.30	#####	657.28	0.00	13.80	84.80	1.40
—	—	—	—	—	—	#####	873.24	18.78	50.20	47.40	2.40
31.25	0.00	0.00	—	—	—	—	—	—	—	—	—
36.06	#####	2.00	8.70	112.25	4.05	#####	2967.14	42.25	104.20	387.00	8.80

probably 1.75 perhaps 3.85
perhaps 1.56 indeed 3.55
indeed 1.13 clearly 0.9

AP1994 vs AP2014									
Group	A (raw)	E (raw)	N (raw)	Total	χ^2	df	p-value	Cramér's V	Effect
AP1994	577	###	32	2500	#####	2	< .001	#####	small
AP2014	174	###	81	2500					

EL1994 vs EL2014									
Group	A (raw)	E (raw)	N (raw)	Total	χ^2	df	p-value	Cramér's V	Effect
EL1994	101	632	9	742	20.997	2	< .001	#####	small
EL2014	521	###	44	2500					

AP1994 vs EL1994									
Group	A (raw)	E (raw)	N (raw)	Total	χ^2	df	p-value	Cramér's V	Effect
AP1994	577	###	32	2500	31.232	2	< .001	#####	small
EL1994	101	632	9	742					

AP2014 vs EL2014									
Group	A (raw)	E (raw)	N (raw)	Total	χ^2	df	p-value	Cramér's V	Effect
AP2014	174	###	81	2500	#####	2	< .001	#####	small
EL2014	521	###	44	2500					

distribution Across C	
Corpus	Epis Attitu No-ξ Total
AP1994	189 577 32 (2500
AP2014	224 174 81 (2500
EL1994	632 101 9 (1. 742
EL2014	193 521 44 (2500

* $\chi^2(6, N=8,242) = 300.030, p < .001$ *, Cramér's V = 0.135 (small effect)

Appendix 5

ALL reference-listed Epistemic & Attitudinal Stance Adverbs

EPISTEMIC ADVERBS

Adverb	AP1994 n	AP1994 /1M	AP2014 n	AP2014 /1M	EL1994 n	EL1994 /1M	EL2014 n	EL2014 /1M
As	134,080	8380.00	173,814	8690.70	1,309	6145.54	33,271	6654.20
Like	9,476	592.25	8,804	440.20	489	2295.77	14,022	2804.40
Really	2,193	137.06	2,014	100.70	245	1150.23	10,689	2137.80
Always	4,951	309.44	4,373	218.65	142	666.67	4,009	801.80
Never	3,652	228.25	2,583	129.15	164	769.95	4,132	826.40
Quite	3,891	243.19	2,436	121.80	166	779.34	3,027	605.40
Probably	3,896	243.50	1,969	98.45	164	769.95	2,103	420.60
Actually	2,557	159.81	2,415	120.75	87	408.45	2,808	561.60
Maybe	204	12.75	418	20.90	170	798.12	1,700	340.00
Perhaps	5,403	337.69	3,990	199.50	77	361.50	893	178.60
Of course	3,753	234.56	2,006	100.30	59	277.00	1,246	249.20
According to	3,578	223.63	5,199	259.95	47	220.66	308	61.60
Indeed	4,785	299.06	5,184	259.20	17	79.81	522	104.40
Clearly	4,409	275.56	3,689	184.45	22	103.29	676	135.20
In fact	3,201	200.06	2,438	121.90	44	206.57	567	113.40
Certainly	2,874	179.63	1,593	79.65	35	164.32	982	196.40
Sort of	1,315	82.19	822	41.10	67	314.55	555	111.00
Definitely	122	7.63	201	10.05	15	70.42	2,133	426.60
Obviously	1,332	83.25	666	33.30	38	178.40	710	142.00
Apparently	1,520	95.00	764	38.20	31	145.54	547	109.40
Mainly	1,648	103.00	2,050	102.50	19	89.20	257	51.40
Typically	803	50.19	2,889	144.45	2	9.39	81	16.20
No doubt	982	61.38	379	18.95	15	70.42	240	48.00
Inevitably	870	54.38	558	27.90	0	0.00	53	10.60
Undoubtedly	629	39.31	327	16.35	2	9.39	60	12.00
Roughly	488	30.50	548	27.40	1	4.69	68	13.60
Evidently	427	26.69	202	10.10	1	4.69	20	4.00
In most cases	288	18.00	259	12.95	0	0.00	20	4.00
Without doubt	33	2.06	19	0.95	2	9.39	23	4.60
Wisely	36	2.25	25	1.25	0	0.00	28	5.60
Sensibly	68	4.25	21	1.05	0	0.00	9	1.80
In our view	41	2.56	36	1.80	0	0.00	1	0.20
Disturbingly	21	1.31	13	0.65	0	0.00	5	1.00
From our perspective	0	0.00	7	0.35	0	0.00	0	0.00

Corpus sizes: AP1994 = 16 M | AP2014 = 20 M | EL1994 = 213 K | EL2014 = 5 M words

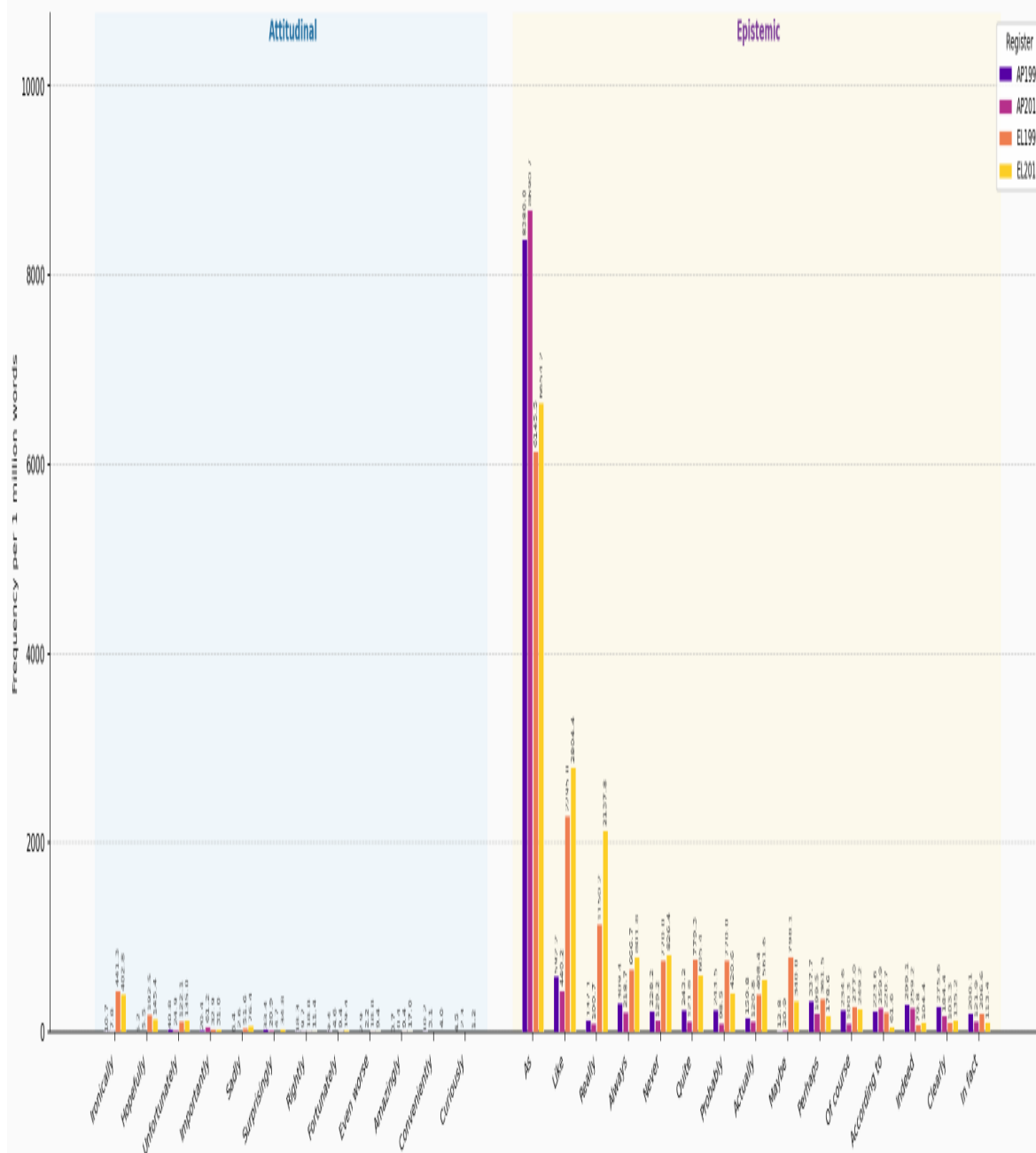
All expressions explicitly mentioned in Biber and Finegan (1988), Biber et al. (1999), and Biber (2006) across corpora

ATTITUDINAL ADVERBS

[illegible]

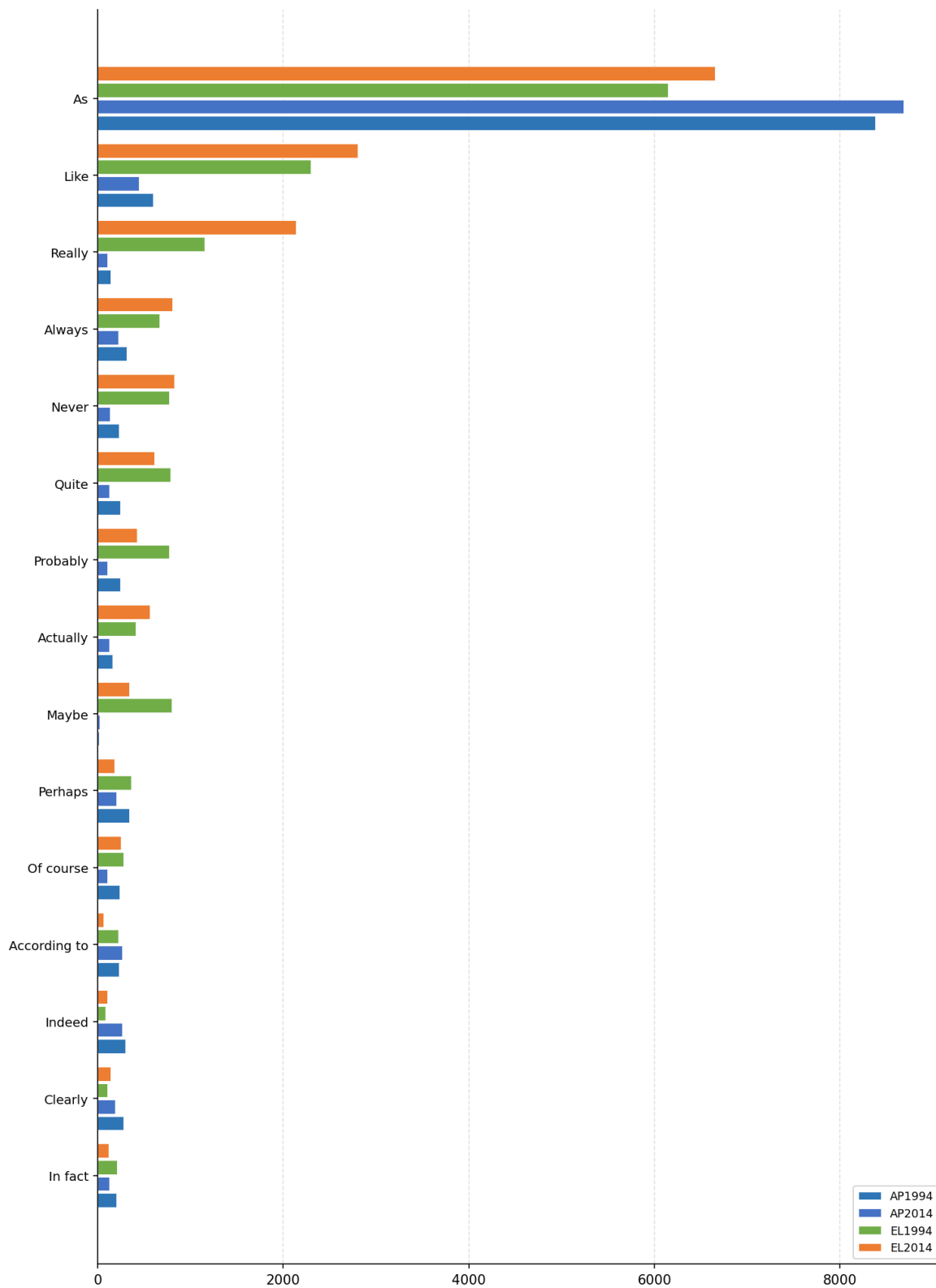
Appendix 5

Stance Adverbial Distribution – Frequency per 1 Million Words



Appendix 5

Top 15 Epistemic Adverbs (per 1 million words)



Appendix 5

Top 12 Attitudinal Adverbs (per 1 million words)

