

EVIDENTIALITY STRATEGIES AS DISTINGUISHING MARKERS IN FORENSIC AUTHORSHIP ANALYSIS

The current research starts from the premise that authors form idiosyncratic habits in how they use evidentiality strategies in academic writing. The goal of the research is to examine whether the frequency of evidentiality strategies is author-specific and if lexico-grammatical expressions used to denote it could be useful distinguishing markers in authorship analysis. Evidentiality is, here, understood in a narrow sense, as a discourse function with a primary meaning to denote source of information. A modification of Gurajac's (2010) classification is used to identify and classify evidentiality. The corpus for this research consists of 19 samples coming from 5 authors, which results in approximately 15,000 words total. The qualitative aspect of the research implies identification and classification of lexico-grammatical expressions used to denote evidentiality, as well as recognizing certain author-specific tendencies in the usage of given expressions. The frequency of evidentiality strategies is expressed per 1000 words and in percentages. One-way analysis of variance is used to test whether between-author variability is higher than within-author variability, that is, whether the tested parameters are useful distinguishing markers in authorship analysis. The results show that the percentage of assumptive inferred evidentiality is rather constant across samples of each author. In addition, there is evidence that the relation of assumptive inference to deductive inference and general knowledge are also idiosyncratic.

Keywords: forensic authorship analysis, evidentiality, evidentiality strategies, inferred evidentiality, reported evidentiality

1. Introduction

Authorship disputes date back from the time of ancient Greece, when playwrights often accused each other of plagiarism. More recently, two major authorship controversies swept the public. In the 18th century, a German priest, H. B. Witter, cast doubt on the authorship of the *Bible*, claiming that "the different names for divinity in the Pentateuch could

* kristinatomic89@hotmail.com

indicate that several authors had contributed to it” (Olsson, 2008: 17). A century later, James Corton Cowell presented the idea that Shakespeare was not the author of all the plays he published, but rather it was Bacon. However, it was not until much later that Svartvik approached authorship analysis methodologically and presented his analysis of the Evans case. This point is usually taken to mark the beginning of forensic linguistics as a science (Olsson, 2008).

With the birth of forensic linguistics came the notion of “*forensic fingerprint*” which implies that every person uses language in a unique way, and the differences in language use can be observed as easily as fingerprints (Coulthard, 2004; Olsson, 2008). Contemporary linguists, however, point to the fact that there is not a final set of features according to which one author can be differentiated from another. Rather, they claim, there is a continuum of possibilities in language use that varies either consciously, depending on the context and function we mean to convey, or subconsciously (Coulthard & Johnson, 2007; Olsson, 2008). It is exactly the subconscious language habits that present potentially viable distinguishing markers in authorship analysis.

In the current paper, our goal is to examine the notion of evidentiality in academic articles and determine whether it could be a suitable forensic marker in author identification. Evidentiality is, here, taken to be a discourse function with a primary meaning of stating the source of information (Plungian, 2001; Aikhenvald, 2003; Gurajac, 2010). The source of information may have several forms. For instance, direct evidentiality implies the author has seen or heard the proposition, or experienced it through other senses. Inferred evidentiality denotes the proposition is reached through author’s belief, assumption or conclusion, while reported evidentiality means the proposition was learned indirectly, from the third party, including known and unknown sources (Gurajac, 2010). It is not uncommon that evidentiality is followed by author’s attitude towards the proposition and/or the source of information, which is in literature referred to as *epistemic modality*. Lexico-grammatical expressions regarding the expectation of knowledge, or *miratives*, may also be present (Aikhenvald, 2003; Gurajac, 2010).

Apart from some research in the context of Italian criminal trials (Greco 2018), the concept of evidentiality has not been explored in

English from the perspective of forensic linguistics. Bearing in mind that, in discourse, linguistic elements are not randomly chosen forms, but rather have their purpose and function in the context (Brown & Yule, 1983); one may be led to conclude that there is little chance of discovering anything speaker-specific regarding evidentiality. However, even when conveying a particular function, authors have a degree of freedom to select certain lexico-grammatical items over others. Furthermore, there might be differences between authors in how frequently they employ particular evidentiality strategies over others. This is exactly what the current research aims to explore.

2. Theoretical background

2.1. Forensic authorship analysis

The basic premises of authorship analysis and attribution is the hypothesis that every person has their own, unique, style of writing or idiolect (Coulthard, 2004; Coulthard & Johnson, 2007). When people use language, either spoken or written, they have a habit of selecting certain lexico-grammatical structures more often than others, as well as combining them in a particular way (Coulthard, 2004; Coulthard & Johnson, 2007). Such habits are said to be author-specific features of language use, and they are often applied in forensic linguistic analysis to identify an author of a questioned text (Coulthard, 2004; Coulthard & Johnson 2007).

A forensic linguistic marker is an aspect of language, a variable, which can be measured and compared across samples (McMenamin, 2002; Rose, 2002; Olsson, 2008). Forensically significant markers exhibit lower within-author than between-author variation, are resistant to disguise and occur relatively frequently in the tested material (Rose, 2002). To date, there does not exist a final list of distinguishing markers that could identify every single author (McMenamin, 2002; Coulthard & Johnson 2007). Rather, a forensic linguistic marker will more easily identify authors whose variable values are towards the edge of the Gaussian curve than those whose values are towards the centre. In other words, it is up to linguists to determine how the disputed text deviates from the average and which linguistic features are potentially useful markers in that particular context (Rose, 2002).

Some of the methods that have been used in the past years to confirm or dispute authorship of a text include stylometrics, corpus analysis, (Coulthard, 2004) and linguistic analysis (Coulthard & Johnson 2007). Stylometrics is primarily concerned with measuring technical aspects of language, such as word or sentence length, lexical richness, pace, hapax legomena (vocabulary items used only once in the text) and so on (McMenamin, 2002; Coulthard & Johnson 2007). Corpus analysis, on the other hand, involves searching for the frequency of certain vocabulary items or strings of vocabulary from the disputed text in a large corpus in order to determine the likelihood of such strings being used by other authors (Coulthard, 2004; Coulthard & Johnson 2007). Finally, linguistic analysis may involve analysing various aspects of idiosyncratic language use, including syntax, grammar, errors and so on (Coulthard & Johnson 2007).

Discourse analysis has proven quite a useful asset in forensic linguistics too (Shuy, 2001; McMenamin, 2002). In legal cases, language experts have relied on the use of discourse markers, patterns of interruption, patterns of register shifting and other aspects of discourse (Shuy, 2001). Shuy (2001) points to the importance of discourse analysis in forensic linguistics and invites linguists to explore the uses of its various aspects for the legal purposes.

2.2. Evidentiality strategies

Evidentiality as a concept is rather controversial, in the sense that there are debates with regard to its definition and scope. One of the first authors describing evidentiality was Boas (1911) who identified it as a semantic category relating to the source of knowledge or source of information (Dendale & Tasmowski 2001). In the broad sense, the term *evidentiality* is used to refer to the source of information as well as the reliability of knowledge or speaker's epistemic values towards the information (Chafe, 1986; Mithuan, 1986). According to this view, the distinction between evidentiality and epistemic modality does not exist and the former encompasses the latter (Dendale & Tasmowski 2001). In the narrow sense, there is a "disjunction between the conceptual domains of evidentiality and [epistemic] modality" (Dendale & Tasmowski 2001: 342). Here, evidentiality is used to refer to stating or implying the existence

of source of information, while epistemic modality is related to speaker's attitude towards the veracity of the proposition, reliability of the source, or even attitude towards the expectation of knowledge (Willet, 1988; Faller, 2006; Aikhenvald, 2014).

More recently, Aikhenvald points to the importance of differentiating between evidentiality systems and evidentiality strategies (Aikhenvald, 2003; 2014). Namely, *evidentiality system* is a term used to describe an obligatory grammatical category of some languages (on a par with gender, cases, tenses etc.), whose primary meaning is denotation of the source of information but may also include modality regarding the factuality of proposition, reliability of the source or expectation of knowledge (Aikhenvald, 2003; 2014). *Evidentiality strategies*, on the other hand, include lexico-grammatical means (modal verbs, conditional / potential mood, perfective aspect), which do not have the source of information as a primary denotation, but may still have some extended evidential-like meaning or gain such meaning in the context (Aikhenvald, 2003).

Bearing in mind the definitions above, it is clear that when we talk about evidentiality in English, we refer to evidentiality strategies rather than evidentiality systems. In the current research, we take a view of evidentiality in the narrow sense, where source of information may be expressed independently of epistemic and other modality. We will use the term *evidentiality strategies* to describe lexico-grammatical means with a discourse function of stating or implying the existence of the source of information. In order to limit the scope of the research, lexico-grammatical means used to give epistemic evaluation regarding the veracity of the proposition, reliability of the source and expectation of knowledge will not be observed in this paper.

2.3. Source of information, epistemic modality and mirativity

There is no universal classification of source of information in literature. This is mostly, as Wierzbicka (1994) points out, due to the fact that linguists observe languages ethnocentrically or Anglocentrically. Depending on the literature, and understanding of the concept of evidentiality, several classifications of 'source of information' can be used as analysis frameworks –Plungian (2001: 354), Gurajak (2010: 72),

Popović (2010: 45), Aikhenvald (2014: 12), to name a few. Wierzbicka (1994), talking primarily about evidentiality systems, insists that if linguists are indeed looking for a universal framework of analysis of evidentials, they cannot rely on an arbitrary classification, since certain aspects of evidentiality present in one language may not necessarily be present in another (Wierzbicka, 1994). Instead, she suggests, evidential expressions should be reduced to universal semantic primitives or near-primitives, as they can provide “constant and language-independent points of reference” (Wierzbicka, 1994: 81).

Below, motivated by Wierzbicka’s warning not to leave out some of the meanings evidentiality can convey, we offer Gurajac’s (2010) classification of the source of information with a few modifications advocated in other literature. Namely, ‘reported evidentiality’ category is divided into three subcategories to include self-quotation (‘autocitiranje’, added by Popović 2010). Similarly, the category ‘inferred evidentiality’ is modified to include ‘general’ knowledge, that Wierzbicka (1994: 84) describes with the primitive “I know P because everyone knows P”, and which is included as part of ‘inference’ category by both Popović (2010) and Aikhenvald (2014).

The proposed framework of source of information classification is the following¹:

A. DIRECT EVIDENTIALITY (including direct access to P, P may be heard, smelled, tasted, etc.)

Visual: ‘P, and I see/saw P.’ (e.g. *We see from (1) that there is a principle of language use that can be stated very informally as (2).*)

Auditory: ‘P, and I hear(d) P.’

Tactile: ‘P, and I feel/felt P.’

Olfactory: ‘P, and I taste(d) P.’

Gustatory: ‘P, and I smell(ed) P.’

B. INFERRED EVIDENTIALITY (evaluative, presenting the speaker’s judgement related to P)

¹ The examples offered in this classification come from the corpus of the current research.

Deductive: '*P, because I can observe a resultant state of a situation or event*', (e.g. *...there must be a rich innate language-specific universal grammar...*)

Assumptive: '*P, because I assume P based on my knowledge and experience*', (e.g. *There is a very good reason to believe that...*)

General knowledge: '*P, because everyone knows P*', (e.g. *It is widely agreed that aspects of linguistic structure can be...*).

C. REPORTED EVIDENTIALITY (non-sensory evidence obtained not through senses or reflection, but from other people or sources)

Quotative: '*P, because I was told that P and I know where P comes from*', (e.g. *Chomsky concluded that there was no reason to consider derivational constraints as a single unified phenomenon*)

Hearsay: '*P, because I was told that P and I do not know where P comes from*', (e.g. *...claims that music is an adaptation favored by natural selection are considerably weaker than those for language..*)

Self-quotation: '*P, because I said P before T_0* ', (e.g. *As was mentioned above, some of the phenomena involving global rules are susceptible to ad hoc solutions.*)

(after Gurajac 2010: 72)

As mentioned in 2.2, modality is, in this paper, viewed as a distinct category in relation to evidentiality (Willet, 1988; Aikhenvald, 2014). Across literature, authors point to different modal meanings an evidential expression can assume. Epistemic modality is understood as “speaker’s degree of certainty and/or the necessity/possibility of the truth of the propositional content”(Faller 2006: 2). Aikhenvald (2014: 14) uses the term ‘dubitative’ to express the meaning of lower degree of certainty regarding the reliability of the source or factuality of the proposition, as in *They allegedly said he is a killer*. In addition, many verbs of reporting in English are semantically coloured and usage of different verbs may be indicative of different degree of certainty regarding the proposition or the speaker’s attitude towards it. The examples *He might be a killer* and *He is thought to be a killer* imply a lower degree of certainty than *He must be a killer* and *He is claimed to be a killer*. Mithun (1986: 89-90) makes a difference between ‘precision of the truth’ and ‘probability of the truth’,

where the first involves lexical expressions and hedges as *sort of, kind of, almost, completely* and the second involves modal verbs, adverbs and adjectives such as *may, might, maybe, probably* etc.

Another category often closely related to evidentiality is what DeLancey refers to as ‘mirativity’ or ‘admirativity’ (DeLancey, 2001: 369). DeLancey (2001: 369) defines mirativity as “the marking of a proposition as representing information which is new to the speaker”. While in some literature, it has been described as a category embedded in evidentiality (see Mithun 1986), DeLancey gives an example to prove that mirativity may exist independently. Namely, in English, we often use rising intonation to signal surprise by the new information, as in *Yeah?* but not the source (DeLancey, 2001: 377). Gilmour, Gonzales and Louie (2010) draw parallels between irony and mirativity, posing a hypothesis that both meanings are “parasitic” on evidentials and both arise from the difference between illocutionary and proposition level. The difference is that for irony to be felicitous, P does not match the reality, while in mirativity, P matches the reality² (Gilmour, Gonzales and Louie 2010: 78). Similarly, Peterson (2010) compares mirativity to metaphor, as both are the result of implicature and flaunting of Gricean maxims (Peterson 2010: 133-134). He gives the following example:

He must be blind! (P = He is blind) [said by a man who is watching a sport when a player misses an excellent opportunity for a score]

Peterson (2010) explains that if the speaker does not know if P is true, this is a regular evidential, or what Gurajac (2010) would call *assumptive inference*. If the speaker knows P is true, then he is flaunting the quantity maxim and expressing mirative meaning, while if the speaker knows P is not true, he is flaunting the quality maxim and expressing nonliteral or metaphoric meaning (Peterson, 2010: 133).

In order to limit the scope of the research, in this paper, we will not deal with epistemic modality and mirativity. Bearing in mind that mirativity and modality may be observed in expressions that do not necessarily denote

2 For example, a sister who sees her brother playing games, while he told her he will be doing his homework: *I see you're doing your homework* (irony). *I see you're not doing your homework* (mirativity).

source of information, the author is of the view that exploration of these concepts is more suitable for studies that take evidentiality in broad sense.

3. Methodology

The goal of the current research is to test whether the choice of evidentiality strategies is author specific, that is, if evidentiality could be used as a marker in forensic authorship identification. Relying on the modification of Gurajac's classification of source of information presented above, we aim to identify lexico-grammatical expressions used to denote evidential meanings in the context.

3.1 Corpus

The corpus consists of 19 samples coming from five authors, famous linguists who have published in the field of cognitive, functional and generative linguistics.³ The samples were selected from scientific papers of the chosen authors in such a way to cover a larger span of time in order to account for within-author variability. Each sample consists of the article introduction and conclusion, although the sections in the papers were sometimes labelled differently (*Issues*, *Final remarks* etc.). The corpus is approximately 15,000 words long, the average length of samples being 774.89 words, with standard deviation of 329.52. While it is probably true that analysing longer samples would provide a clearer picture of the use of evidentials by individual authors, in forensic reality, linguists most often deal with quite short texts, even under 200 words (Coulthard & Johnson, 2007: 162). Therefore, the more robust a parameter is across short texts, the higher value it has in forensic linguistics.

3.2. Procedure

Looking for consistency and resemblance (McMenamin, 2002) in the use of evidentiality within one author's work and between authors, we performed quantitative and qualitative analysis of the corpus. The frequency of occurrence of evidentiality is expressed per 1000 words in

³ In this paper, the authors will be referred to as A1, A2, A3, A4, and A5. To see the complete list of authors and papers used as the corpus in this research, refer to the section *Sources* at the end.

order to account for the variability of the sample size. Such normalized scores are required for better comparability of the results as well as for statistical analysis. To observe the percentage of individual evidentiality categories in relation to all evidentiality found in a text, we also expressed the results per 100 evidentials. The difference between authors and within-author variability for each parameter are determined with a single factor analysis of variance (ANOVA), while predominance of the presence of one evidentiality strategy over another is expressed through z-scores. The qualitative analysis focuses on determining whether authors exhibit any differences in the choice of lexico-grammatical expressions used to denote evidentiality.

4. Results

Identifying evidentiality in a text is not a straightforward task, as one cannot solely rely on the form of expressions or the semantic meaning of lexemes that appear in the text. For example, a verb of perception *see* can be used to convey both direct - visual and inferred - deductive evidentiality. For example, saying *We see that in this sentence the verb takes on a suffix* implies direct evidentiality, while *I see what you mean* implies deduction. Similarly, we cannot rely solely on grammatical form to determine the evidentiality category. For example, the passive voice in *It is said that he has left* marks the hearsay evidentiality, while the passive with the agent *It is said by modern linguists, such as Pinker and Jackendoff, that...* is reported evidentiality. Therefore, evidential meaning of expressions can only be determined in the context of the discourse.

4.1. Quantitative analysis

In the nineteen samples, we identified the total of 162 expressions of evidentiality, 6 sensory, 113 inference and 43 reported evidentials. On average, this translates to approximately 10.8 evidential expressions per 1000 words. In the given section, we calculate the frequency of evidentiality as used by each of the five authors and determine if its patterns of usage are speaker-specific and to what extent.

4.1.1. Source of information

In the table below, we can observe the average frequency and standard deviation of evidentiality in total, as well as of individual evidentiality categories per author.

Table 1. Evidentiality strategies - denotation of source of information; expressed per 1000 words with standard deviation for individual authors

Evidentiality	Author 1		Author 2		Author 3		Author 4		Author 5	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Direct	1.58	3.15	0	0	0	0	0.81	1.63	0	0
Visual	1.58	3.15	0	0	0	0	0.81	1.63	0	0
Inferred	7.90	3.80	6.78	1.53	7.48	3.42	9.16	2.99	5.89	4.25
Assumptive	0.53	1.05	0.37	0.74	1.64	1.90	1.75	1.22	3.16	1.65
Deductive	4.61	3.09	4.75	0.38	5.45	1.87	7.16	3.79	1.49	1.50
General Knowledge	2.76	2.82	1.66	2.08	0.39	0.77	0.25	0.50	1.24	1.14
Reported	3.84	6.51	2.61	2.88	1.53	2.17	3.78	3.10	2.12	1.21
Quotative	3.84	6.51	1.47	2.10	1.53	2.17	3.53	2.89	0.74	0.75
Hearsay	0	0	0.57	1.15	0	0	0.25	0.50	1.38	1.31
Self-quotation	0	0	0.56	1.11	0	0	0	0	0	0
Total	13.31	3.50	9.57	3.78	9.01	4.16	13.76	1.71	8.01	4.41

The results confirmed that all of the authors relied on evidentiality to some extent; on average we found between 8 and 14 evidentiality expressions per 1000 words of text. The lowest variance is noted for A4, while A3 and A5 exhibit rather high variance in the usage of evidentiality across samples. Only two out of five authors employ direct evidentiality; however, since it was not present in every sample by these authors, we cannot claim that it is an author-specific phenomenon. Inferred evidentiality appears to be

more common than direct or reported evidentiality; on average, authors use between 6 and 9 inferred evidentiality expressions per 1000 words. With regard to inferred evidentiality, the author with the lowest variance is A2, while A5 and A1 exhibit high within-author variability. Majority of the authors (A1, A2, A3 and A4), seem to rely on deductive inference more frequently than on assumptive inference or general knowledge. The exception is A5, who employs assumption slightly more often than the other two inferred evidentiality sub-categories. With regard to reported evidentiality, it appears rather irregularly, the most common form being quotative. Hearsay evidentiality was found only with three authors (A2, A4 and A5) on one or two occasions, while self-quotation was employed only by A2. The presence of reported evidentiality appears to depend on the nature of scientific paper: whether the authors present some new theory or view, or propose reshaping of some older ideas and opinions.

The given results were subject to the statistical test; however, it did not confirm significant differences between authors or provide reason to regard any of the evidentiality categories as potential author-specific parameter when expressed per 1000 words (see Table 2).

Table 2. Evidentiality strategies - denotation of source of information; one-way ANOVA, significant for $p < 0.05$

Evidentiality	Between-author			Within-author			F-ratio	p-value
	SS	df	MS	SS	df	MS		
Direct	7.77	4	1.94	37.75	14	2.70	0.721	0.5919
Visual	7.77	4	1.94	37.75	14	2.70	0.721	0.5919
Inferred	21.51	4	5.38	148.43	14	10.60	0.507	0.7314
Assumptive	17.35	4	4.34	25.67	14	1.83	2.365	0.1030
Deductive	56.94	4	14.24	87.14	14	6.22	2.287	0.1115
General Knowledge	16.75	4	4.19	41.98	14	3.00	1.397	0.2856
Reported	16.06	4	4.01	197.83	14	14.13	0.284	0.8834
Quotative	27.80	4	6.95	180.61	14	12.90	0.539	0.7098
Hearsay	4.36	4	1.09	8.22	14	0.59	1.856	0.1744
Self-quotation	0.98	4	0.24	3.72	14	0.27	0.921	0.4791
Total	102.16	4	25.54	179.19	14	12.80	1.995	0.1506

Table 2 shows the results of one-way analysis of variance for the frequency of evidentiality in total, as well as the frequency of individual evidentiality categories. With p-values higher than 0.05, there is no evidence to confirm the significance of difference in the usage of evidentiality between authors. Since within-author variability is higher than between-author variability, the strategies cannot be taken as author-specific parameters in forensic author identification.

Furthermore, we went on to test whether authors differ in the percentage of individual evidentiality categories (calculated per 100 evidentials).

Table 3. Average percentage of evidentiality strategies with standard deviation (per 100 evidentials)

Evidentiality	Author 1		Author 2		Author 3		Author 4		Author 5	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Direct	9.38	19.75	0	0	0	0	5	10	0	0
Inferred	65.03	37.15	77.02	22.26	86.51	20.27	65.96	16.29	66.60	29.59
Assumptive	3.13	6.25	2.63	5.26	18.05	23.73	12.54	9.04	39.54	9.11
Deductive	41.42	36.05	57.42	28.89	65.67	23.09	51.50	25.81	14.51	12.69
Gen. know.	20.49	17.91	16.96	20.90	2.78	5.56	1.93	3.85	12.55	10.93
Reported	25.59	40.84	21.66	19.96	13.49	20.27	29.04	23.50	33.40	29.59
Quotative	25.59	40.84	11.47	15.00	13.49	20.27	27.12	21.90	7.25	6.34
Self-quotat.	0	0	3.95	7.89	0	0	0	0	0	0
Hearsay	0	0	6.25	12.5	0	0	1.92	3.85	26.14	35.58

Table 4. Average percentage of evidentiality strategies (per 100 evidentials); one-way ANOVA, significant for $p < 0.05$

Evidentiality	Between-author			Within-author			F-ratio	p-value
	SS	df	MS	SS	df	MS		
Direct	251.16	4	62.79	1375.46	14	98.25	0.639	0.643
Inferred	1365.24	4	341.31	9405.68	14	671.83	0.508	0.731
Assumptive	3021.43	4	755.36	2301.05	14	164.36	4.596	0.014

Deductive	5190.95	4	1297.74	10322.1	14	737.29	1.760	0.193
Gen. know.	1109.02	4	277.25	2648.64	14	189.19	1.465	0.265
Reported	839.64	4	209.91	10840	14	774.34	0.271	0.892
Quotative	1156.72	4	289.18	8430.54	14	602.18	0.480	0.750
Self-quotat.	49.21	4	12.30	186.98	14	13.36	0.921	0.479
Hearsay	1571.62	4	392.90	3045.49	14	217.53	1.806	0.184

On average, direct evidentiality takes as much as 9.38% of all evidentiality strategies for A1 and 5% for A4, while A3 and A5 do not employ this form of evidentiality at all (Table 3). Ranging between 65 and 87 per cent, inferred evidentiality accounts for the majority of evidentiality strategies for each author, while reported evidentiality on average ranges between 13 and 33 per cent. None of the evidentiality categories exhibit significant between-author differences, as within-author variance is rather high in all cases (Table 4).

Observing individual sub-categories, however, we may note that the percentage of assumptive inference is rather speaker specific. The between-author variability is significantly higher than within-author variability (Table 4). On average, for some authors (A1 and A3), assumptive inference makes for only about 3% of all evidentiality strategies, while for others (such as A5) it accounts for as much as 40% of all evidentiality (Table 3). As such, the percentage of assumptive inference can be identified as a useful marker in forensic authorship analysis.

4.1.2. Inferred evidentiality extended

Having confirmed that the inferred evidentiality makes for most of evidentiality strategies for each of the five authors, as well as that the percentage of assumptive inference is speaker-specific, we decided to take a closer look at inferred evidentiality sub-categories and their relations.

Table 5. Inferred evidentiality - average percentage of inference sub-categories with standard deviation

Inferred evidentiality	Author 1		Author 2		Author 3		Author 4		Author 5	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Assumptive	5	10	5.56	11.11	18.75	23.94	22.02	18.17	67.24	28.76
Deductive	67.22	35.71	73.61	20.97	78.13	25.77	74.41	21.06	17.66	15.88
Gen. know.	27.78	27.31	20.83	25.00	3.12	6.25	3.57	7.14	15.10	13.08
Total	100%		100%		100%		100%		100%	

Table 6. Inferred evidentiality - average percentage of inference sub-categories; one-way ANOVA, significant for $p < 0.05$

Inferred evidentiality	Between-author			Within-author			F-ratio	p-value
	SS	df	MS	SS	df	MS		
Assumptive	8412.61	4	2103.15	5034.14	14	359.58	5.849	0.0055
Deductive	8077.51	4	2019.38	8972.88	14	640.92	3.151	0.0482
Gen. know.	1857.78	4	464.45	4724.65	14	337.48	1.376	0.2921

As the p-values in Table 6 indicate, both the percentage of assumptive and of deductive inference in relation to the total number of inferred evidentials seem to exhibit significant differences across authors. In other words, while authors do not differ in the number of inferred evidentials per 1000 words, the percentage of individual sub-categories under inferred evidentiality tends to be rather author-specific.

With this in mind, we went on to examine the relation of assumptive, deductive and general knowledge evidentiality to determine if some authors use one sub-category more often than other, and how authors differ in this regard. To do this, first, we calculated the z-scores of each of the two sub-strategies in relation to the number of words, and then to the number of evidential expressions found in the corpus. Table 7 below provides average z-scores for each of the two sub-categories per author.

Table 7. Inferred evidentiality - average z-scores of inference sub-categories for each author; expressed per number of words and number of evidentials

Relation	compared to	Author 1	Author 2	Author 3	Author 4	Author 5
assumptive to deductive	words	-1.39238	-1.55878	-1.13898	-1.37958	0.9231
	evidentials	-1.7081	-1.8737	-1.4542	-1.73268	1.08487
assumptive to general knowledge	words	-0.85448	-0.32995	0.5779	0.93153	1.0458
	evidentials	-0.93323	-0.3925	0.65813	0.97768	1.22267
deductive to general knowledge	words	0.69643	1.16085	1.69733	2.13028	0.12617
	evidentials	0.94208	1.4149	2.0916	2.527	0.14137

Table 8. Inferred evidentiality - average z-scores of inference sub-categories for each author, expressed per number of words and number of evidentials; one-way ANOVA, significant for $p < 0.05$

Relation	compared to	Between-author SS	Within-author SS	F-ratio	p-value
assumptive to deductive	words	13,613	9,216	5,16989	0,00904
	evidentials	19,849	15,950	4,35553	0,01701
assumptive to general knowledge	words	10,408	10,547	3,45388	0,03663
	evidentials	12,873	12,541	3,59243	0,03240
deductive to general knowledge	words	8,925	11,712	2,66728	0,07636
	evidentials	12,429	18,437	2,35942	0,10361

As indicated in Table 7, the difference between speakers regarding average z-scores is significant for the first two markers. That is, when discussing relation of assumptive and deductive inferred evidentiality, we can observe that the first four authors (A1, A2, A3 and A4) predominantly rely on deductive sub-category (which is indicated by negative z-scores in Table 7), while A5 consistently uses more assumption than deduction in his papers. Similarly, we may note that A1 and A2 rather rely on general knowledge as source of information while A3, A4 and A5 prefer to use

assumptive inference. Therefore, we may conclude that in scientific papers, the relation of the number of assumptive evidentiality occurrences to other inference sub-categories may be indicative of an author's style of writing.

4.2. Qualitative analysis

Authors use various lexico-grammatical means to denote source of information. They may rely on verbs of belief or opinion, reporting verbs and clauses, parentheticals or some grammatical categories such as conditionals. Our goal is to determine whether certain authors have a preference for particular lexemes and forms, and if they could be identified on the basis of these expressions. Below, we will examine the lexico-grammatical expressions found in the corpus for every evidentiality category and sub-category and try to establish if any of them are author-specific.

4.2.1. Direct evidentiality

The only form of direct evidentiality that was noted in the corpus is visual. It was employed by only two authors (A1 and A4), both of whom used three expressions in only one of the analysed samples. The verb of perception *see* appears in all examples.

- (1) *We see this in the centrality of full argument structure to grammar.*
- (2) *We see it in the fact ...*
- (3) *And we see it in the limited place for the drive to reduce ambiguity as a functional force affecting language.*
- (4) *We see from (1) that there is a principle of language use that can be stated very informally as (2).*
- (5) *...more complex possibilities exist, as we will see presently.*
- (6) *... first the plot thickens considerably, as we now see.*

Even though both authors use the same verb to denote the visual source of information, there is some difference in the form. Namely, in (1), (2) and (3) coming from A1, deictic expressions *this* and *it* are used to refer to the same proposition the author had mentioned before, while in (4), (5) and (6), the author (A4) includes the proposition in the same sentence with the verb *see*.

4.2.2. Inferred evidentiality

Inferred evidentiality can be assumptive, where the author expresses the proposition as his/her belief or assumption based on his/her knowledge and experience; deductive, where the author arrives at the proposition through the process of reasoning or conclusion and general knowledge, where the author presents the proposition as something well-known and factual. In appendices I, II and III, we can see some of the examples of inferred evidentiality found in the corpus.

Assumptive evidentiality in the given corpus is expressed through verbs of attitude and belief such as *argue, believe, claim, consider, deny, doubt, hold, maintain, suspect, think*. As shown in I, Author 5 employs assumptive evidentiality considerably more than other authors. In fact, he is the only one to use the verb *claim* to express his attitude towards the proposition. In addition, this author has a tendency to employ negation with verbs of opinion/belief or reporting verbs. Such a tendency may be described as a marker of his style.

- (7) *I claim that nouns and verbs also lend themselves to schematic semantic characterization*
- (8) *I do not claim neutrality or deny that important empirical, theoretical, and methodological issues are at stake...*
- (9) *First, I do not hold that all grammatical classes are strictly definable in notional terms:*

With regard to deductive evidentiality, verbs *appear* and *seem* as well as discourse markers *thus, therefore* etc. are quite common in the corpus. A3 uses the copulative verb *appear* with the verb to be in 3 out of 4 samples, however, as three more authors use this verb in at least one of their papers, we should not observe it as an author-specific feature. The same is true of the verb *seem*, which is predominantly used by A4, but appears in the papers by other authors as well. A5 once again emerges as an author with a unique manner of expression. Namely, the expression *It/This is reasonable...* was found in two out of three analysed samples by this author.

- (10) *There seems to be less vehement dispute about the parallel issue in music cognition and acquisition.*

- (11) *...it appears that there is not a unified solution to what appears superficially a unified class of processes.*
- (12) *It is both reasonable and economical to identify the notions subject and object as its clause-level manifestations*

Finally, expressions used to denote that the proposition is part of general knowledge often include nouns *fact* or *view* with a complement, as well as *It* clauses as in *It is agreed that* or *It is well known that*. To express the ideas held by linguists in a particular period of time, A2 uses the passive voice of the verb *see* (*was/were seen*). However, it is rather difficult to make any generalizations regarding the authors' tendencies in the usage of evidential expressions, especially since same or similar expressions tend to be used a few times in one sample, but not necessarily across samples.

- (13) *It is widely agreed that aspects of linguistic structure can be related to various cognitive and communicative functions.*
- (14) *Traditionally dominant has been the view that a category is defined by a set of CRITERIAL ATTRIBUTES,*

4.2.3. Reported evidentiality

Reported evidentiality conveys the meaning that the information/proposition is known through an indirect source. The source of information may be known, which is the case in quotative evidentiality, or unknown, which is labelled hearsay evidentiality. If the author refers to what he/she has already said, that is an example of self-quotation. The only example of self-quotation was noted for A2:

- (15) *As was mentioned above, some of the phenomena involving global rules are susceptible to ad hoc solutions.*

Hearsay evidentiality in the corpus was realized through impersonal passive expressions, such as *are/was claimed to be*, *are considered*, *is deemed*, and *were assumed to be*. There is one instance of a noun with a complement as in:

- (16) *...claims that music is an adaptation favored by natural selection are considerably weaker than those for language.*

Most of the hearsay evidentiality comes from A5, and it appears in 2 out of 3 analysed samples.

Quotative evidentiality is the most common form of reported evidentiality found in the corpus. The source of information is usually animate (such as an author who is usually mentioned by his/her surname), however, there are a few examples of the source being an inanimate entity or an entire school of thought, such as *research*, *examination*, *paper*, *prospectus*, *Gestalt school* etc. Quotative evidentiality usually takes form of an active reporting clause as *He argues that (...)*, *He insists that (...)*; however, there are some examples of the passive voice too: *it has been advocated by*, *it is claimed by* etc. It is far less common to encounter noun phrases as: *Pinker's hypothesis that*, *Patel's view that* or a prepositional phrase *In the view of Heine et al (...)*.

Lexical items used to convey quotative evidentiality are miscellaneous; verbs with a reporting function that can be found in the corpus include *advocate*, *argue*, *believe*, *claim*, *conclude*, *determine*, *emphasize*, *feel*, *insist*, *mention*, *think*, *point out*, *point to*, *propose*, *sketch*, *state*, *stress* and *suggest*. Bearing in mind that most of these verbs appear only once in the corpus with a quotative evidential function, we may safely conclude that all of the five authors whose work we analysed have rather rich vocabulary. Thus, it is quite difficult to make generalizations regarding author-specific patterns. It is however, worth mentioning that A3 is the only one to repeatedly use the verb *show* to denote quotative evidentiality with an inanimate subject.

- (17) *An examination of signed language shows that its structural representation of space systematically differs from that in spoken language*
- (18) *... this paper has shown that there is psychological reality to a certain fundamental conceptual entity with possibly universal linguistic expression*

Furthermore, it is peculiar that A4 uses the verb *conclude* only in the present simple tense (11) and (12), while A2 uses it in the past simple tense (13).

- (19) *Patel (2008), surveying much the same evidence as I have here, concludes the glass is half full rather than half (or three-quarters) empty.*
- (20) *He concludes (Fodor 1980a: 71): ... truth, reference, and the rest of the semantic notions aren't psychological categories.*
- (21) *Chomsky concluded that there was no reason to consider derivational constraints as a single unified phenomenon.*

4. Discussion and Conclusion

The current research started from the premise that authors form idiosyncratic habits in how they use evidentiality in academic papers. The goal was to examine whether the frequency of evidentiality strategies is author-specific and if lexico-grammatical expressions used to denote it could be useful distinguishing markers in authorship analysis. Evidentiality is, here, understood in a narrow sense, as a discourse function with a primary meaning to denote source of information.

With regard to the frequency of evidentiality strategies, results have shown that within-author differences are significantly higher than between-author differences. Therefore, in this corpus, the frequency of evidentiality strategies cannot be regarded as a reliable marker in authorship analysis. One of the reasons for such a great within-author variability in this research may be the fact that the selected papers span across a large period of time. Thus, it is not unusual that authors express a lot of variability across the analysed samples. Further studies should incorporate samples that were written in a narrower time frame, for instance, up to five or ten years.

On the other hand, the results have shown that authors do tend to differ in the percentage of individual evidentiality categories. In particular, the authors exhibited consistency across samples in the percentage of assumptive inference. In addition, the relation of usage of assumptive inference to deductive inference and general knowledge seems to be rather constant too. In the selected group of authors, A5 emerges as the one whose habits in the use of evidentiality diverge from the average the most. In particular, it is easy to identify his samples because he constantly uses assumptive inference more often than deductive inference, as opposed

to other authors who rely on deductive evidentiality considerably more. The fact that this parameter proved robust to writing across such a large span of time and in such a small corpus implies significance for forensic linguistics.

Observing lexico-grammatical expressions of evidentiality, we noticed certain consistencies across samples for some of the authors. For instance, the choice of a reporting verb may be indicative of an author's preference for a particular lexeme, such as the usage of *claim* by A5 or consistent usage of *show* with an inanimate subject by A3. The same is true for the choice of tenses. For example, A4 uses the reporting verb *conclude* in the present tense, while A2 uses it in the past simple. Comparing the tenses and modal verbs in evidential expressions should be a task of future research, as it may bring more insight into author-specific aspects of evidentiality.

To conclude, while there does not seem to be significant difference across authors in the use of evidentiality on a macro level, we did notice certain author-specific habits in the choice of the type of evidentiality and lexico-grammatical evidential expressions. Such findings could be useful for forensic linguistics, authorship analysis in particular. It is important to point out though, that, just as with any forensic linguistic marker, it is only the authors who diverge from the average the most that can easily be identified, such as in the case of A5 in this research. Finally, we should bear in mind that distinguishing markers for a disputed text must be chosen with regard to the context, thus considering evidentiality strategies may not be appropriate for every kind of forensic text. Future research in this direction should be concerned with examining author-specific aspects of evidentiality in different genres and in different languages.

REFERENCES

- Aikhenvald, A. Y. (2003). Evidentiality in typological perspective. In A. Y. Aikhenvald, & R. W. Dixon (Eds.), *Studies in Evidentiality* (pp. 1-31). Amsterdam - The Netherlands: John Benjamins Publishing Company.
- Aikhenvald, A. Y. (2014). The grammar of knowledge: a cross-linguistic view of evidentials, and the expression of information source. In A. Y. Aikhenvald,

- & R. W. Dixon (Eds.), *The Grammar of Knowledge: A Cross-Linguistic Typology* (pp. 1-82). Oxford: Oxford University Press.
- Boaz, F. (1911). Kwakiutl. In F. Boaz (Ed.), *Handbook of American Indian Languages. Part I* (pp. 423-557). Washington, DC: Government Printing Office.
- Brown, G., & Yule, G. (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.
- Chafe, W. (1986). Evidentiality in English Conversation and Academic Writing. In W. Chafe, & J. Nichols (Eds.), *Evidentiality: The Linguistic Coding of Epistemology* (pp. 261-272). Norwood, New Jersey: Ablex Publishing Corporation.
- Coulthard, M. (2004). Author Identification, Idiolect and Linguistic Uniqueness. *Applied Linguistics*, 25(4), 431-447.
- Coulthard, M., & Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. Abingdon, Oxon: Routledge.
- DeLancey, S. (2001). The mirative and evidentiality. *Journal of Pragmatics*, 33, 369-382.
- Dendale Patrick, & Tasmowski, L. (2001). Introduction: Evidentiality and related notions. *Journal of Pragmatics*, 33, 339-348.
- Faller, M. (2006). Evidentiality and Epistemic Modality at the Semantics/Pragmatics Interface. *Paper presented at the University of Michigan Workshop in Philosophy and Linguistics*. Retrieved September 10th, 2018, from <http://www.eecs.umich.edu/~rthomaso/lpw06/fallerpaper.pdf>
- Gilmour, D., Gonzales, A., & Louie, M. (2010). Evidentials and Parasitic Irony: Activating the Illocution-Proposition Distinction. In T. Peterson, & U. Sauerland (Eds.), *Evidence from Evidentials* (Vol. 28, pp. 75-88). Vancouver, Canada: University of British Columbia Working Papers in Linguistics.
- Greco, P. (2018). Evidentiality and epistemic modality in witness testimony in the context of Italian criminal trials. *Journal of Pragmatics*, 128, 128-136.
- Gurajak, B. (2010). Evidentiality in English and Polish. University of Edinburgh. Retrieved September 1st, 2018, from <http://hdl.handle.net/1842/5313>
- McMenamin, G. R. (2002). *Forensic Linguistics: Advances in Forensic Stylistics*. Boca Raton, London, New York, Washington, D.C.: CRC Press.
- Mithun, M. (1986). Evidential diachrony in Northern Iroquoian. In W. Chafe, & J. Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 89-112). Norwood, New Jersey: Ablex Publishing Corporation.

- Olsson, J. (2008). *Forensic Linguistics* (Second edition ed.). London and New York: Continuum International Publishing Group.
- Peterson, T. (2010). Examining the Mirative and Nonliteral Uses of Evidentials. In T. Peterson, & U. Sauerland (Eds.), *Evidence in Evidentials* (Vol. 28, pp. 129-159). Vancouver, Canada: University of British Columbia Working Papers in Linguistics.
- Plungian, V. A. (2001). The place of evidentiality within the universal grammatical space. *Journal of Pragmatics*, 33, 349-357.
- Popović, Lj. (2010). Kategorija evidencijalnosti u srpskom i ukrajinskom jeziku. *ZbornikMatice srpske za slavistiku*, 77, 17-47.
- Rose, P. (2002). *Forensic Speaker Identification*. London and New York: Taylor & Francis.
- Shuy, R. W. (2001). Discourse Analysis in the Legal Context. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The Handbook of Discourse Analysis* (pp. 437-452). Oxford: Blackwell.
- Wierzbicka, A. (1994). Semantics and Epistemology: The Meaning of 'Evidentials' in a cross-linguistic perspective. *Language Sciences*, 16(1), 81-137.
- Willet, T. (1988). A crosslinguistic survey of the grammaticalization of evidentiality. *Studies in Language*, 12, 51-97.

Sources

- Jackendoff, R. (1991). The Problem of Reality. *Noûs. Special Issue on Cognitive Science and Artificial Intelligence*, 25(4), 411-433. Retrieved September 18th, 2018, from <https://www.jstor.org/stable/2216072>
- Jackendoff, R. (1992). Mme. Tussaud Meets the Binding Theory. *Natural Language & Linguistic Theory*, 10(1), 1-31. Retrieved September 18th, 2018, from <https://www.jstor.org/stable/4047765>
- Jackendoff, R. (2008). 'Construction after Construction' and Its Theoretical Challenges. *Language*, 84(1), 8-28. Retrieved September 18th, 2018, from <https://www.jstor.org/stable/40071010>
- Jackendoff, R. (2009). Parallels and Nonparallels between Language and Music. *Music Perception: An Interdisciplinary Journal*, 26(3), 195-204. Retrieved September 18, 2018, from <https://www.jstor.org/stable/10.1525/mp.2009.26.3.195>
- Lakoff, G. (1970). Global Rules. *Language*, 46(3), 627-639. Retrieved September 18th, 2018, from <https://www.jstor.org/stable/412310>

- Lakoff, G. (1993). The Contemporary Theory of Metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 202-251). New York: Cambridge University Press.
- Lakoff, G. (2005). A Cognitive Scientist Looks at Daubert. *American Journal of Public Health. Supplement 1*, 95(S1), S114-S120.
- Lakoff, G. (2012). Explaining Embodied Cognition Results. *Topics in Cognitive Science*, 4, 773-785. doi:10.1111/j.1756-8765.2012.01222.x
- Langacker, R. W. (1986). An Introduction to Cognitive Grammar. *Cognitive Science*, 10, 1-40.
- Langacker, R. W. (1987). Nouns and Verbs. *Language*, 63(1), 53-94. Retrieved September 18th, 2018, from <https://www.jstor.org/stable/415384>
- Langacker, R. W. (1995). Raising and Transparency. *Language*, 71(1), 1-62. Retrieved September 18th, 2018, from <https://www.jstor.org/stable/415962>
- Newmeyer, F. J. (2001). Deconstructing grammaticalization. *Language Sciences*, 23, 187-229.
- Newmeyer, F. J. (2001). The Prague School and North American Functionalist Approaches to Syntax. *Journal of Linguistics*, 37(1), 101-126. Retrieved September 18th, 2018, from <https://www.jstor.org/stable/4176644>
- Newmeyer, F. J. (2002). Optimality and Functionality: A Critique of Functionally-Based Optimality-Theoretic Syntax. *Natural Language & Linguistic Theory*, 20(1), 43-80. Retrieved September 18th, 2018, from <https://www.jstor.org/stable/4048048>
- Newmeyer, F. J. (2003). Grammar Is Grammar and Usage Is Usage. *Language*, 79(4), 682-707. Retrieved September 18th, 2018, from <https://www.jstor.org/stable/4489522>
- Talmy, L. (1988). Force Dynamics in Language and Cognition. *Cognitive Science*, 12, 49-100.
- Talmy, L. (1991). Path to Realization: A Typology of Event Conflation. *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on The Grammar of Event Structure*, (pp. 480-519).
- Talmy, L. (2003). The Representation of Spatial Structure in Spoken and Signed Language: A Neural Model. *Language and Linguistics*, 4(2), 207-250.
- Talmy, L. (n.d.). The Relation of Grammar to Cognition: a Synopsis. *TINLAP '78 Proceedings of the 1978 workshop on Theoretical issues in natural language processing*, (pp. 14-24). Urbana-Champaign, Illinois.

Kristina D. Tomić

STRATEGIJE EVIDENCIJALNOSTI KAO DISTINKTIVNO OBELEŽJE U FORENZIČKOJ ANALIZI AUTORSTVA

Sažetak

Ovo istraživanje kreće od premise da autori imaju idiosinkretičke navike kako koriste strategije evidencijalnosti u akademskom pisanju. Cilj istraživanja je da se ispita da li je frekventnost strategija evidencijalnosti svojstvena autorima i da li se leksičko-gramatički izrazi koji se koriste da iskažu evidencijalnost mogu smatrati distinktivnim obeležjima pri analizi autorstva. Evidencijalnost se ovde definiše u užem smislu, kao diskursna funkcija sa primarnim značenjem izvora informacije. Za klasifikaciju i identifikaciju strategija evidencijalnosti, koristi se Gurajakova podela (Gurajac, 2010) sa malim izmenama. Istraživački korpus se sastoji od oko 15 000 reči koje čini ukupno 19 uzoraka 5 različiti autora, Kvalitativni deo istraživanja podrazumeva identifikaciju i klasifikaciju leksičko-gramatičkih izraza evidencijalnosti, kao i prepoznavanje sklonosti pojedinih autora ka upotrebi određenih izraza. Frekventnost strategija evidencijalnosti iskazana je na 1000 reči i u procentima. Da bi se proverila varijabilnost u okviru uzoraka jednog autora i između autora koristi se jednofaktorska analiza varijanse (ANOVA). Rezultati su pokazali da je procenat inferencijalne evidencijalnosti zasnovane na pretpostavci relativno konstantan u različitim uzorcima teksta jednog autora. Takođe, postoje dokazi da je odnos inferencijalne evidencijalnosti zasnovane na pretpostavci i evidencijalnosti zasnovane na dedukciji i opšteprihvaćenim stavovima takođe idiosinkretičan.

Ključne reči: forenzička analiza autorstva, evidencijalnost, strategije evidencijalnosti, inferencijalna evidencijalnost, raportivna evidencijalnost

Appendices

I - Inferred assumptive evidentiality – lexico-grammatical expressions and examples

	Lexico-grammatical item	Example	Sample	Author
a s s u m p t i v e	argue	<i>I argue that the mental grammar is...</i>	1	A1
		<i>I argue, however, that</i>	1	A4
	believe	<i>However, as it stands, I believe....</i>	2	A3
		<i>I believe that...</i>	1	A5
		<i>...an analysis of raising that I believe to be ...</i>	1	A5
		<i>I believe these advantages are....</i>	1	A5
		<i>There is a very good reason to believe that...</i>	1	A2
	claim	<i>There is a very good reason for believing that...</i>	1	A2
		<i>While I do not claim...that...</i>	1	A5
	consider	<i>I claim that....</i>	3	A5
		<i>I do consider it compatible with....</i>	2	A5
	deny	<i>I do not.... deny that...</i>	2	A5
	doubt	<i>I see little reason to doubt...</i>	3	A5
	hold	<i>First, I do not hold that...</i>	3	A5
	maintain	<i>I maintain that...</i>	3	A5
		<i>I maintain that...</i>	3	A5
	suspect	<i>I suspect...will be to...</i>	4	A4
think	<i>...but I think in fact...</i>	3	A4	
	<i>At least I think this is...</i>	3	A4	
will (modal)	<i>...this set will contain...</i>	4	A3	

II - Inferred deductive evidentiality – lexico-grammatical expressions and examples

	Lexico-grammatical item	Example	Sample	Author
deductive	apparent	<i>It is now apparent that...</i>	2	A3
		<i>It is apparent that...</i>	4	A1
	appear	<i>...in the direction of what appear to be...</i>	1	A3
		<i>There appears to be a ...</i>	3	A3
		<i>...that appears to be...</i>	4	A3
		<i>It appears that...</i>	4	A4
		<i>...what appears superficially a...</i>	4	A4
		<i>Phonologists appear to have...</i>	2	A1
		<i>...grammaticalization appears to be...</i>	3	A1
		<i>...a procedural move that, on the surface of it, appears not to...</i>	2	A2
	clear	<i>It seems clear that...</i>	1	A2
		<i>It has become clear...that...</i>	1	A2
	conclusion	<i>...conclusion that...</i>	2	A4
	conditional clause	<i>...this analysis could never have been arrived at were these not...</i>	1	A5
		<i>This is not surprising, since if they were all of exactly the same form, they would be...</i>	1	A2
		<i>If such criteria were applied to transformational rules, they would...</i>	1	A2
		<i>If one applied Chomsky's argument..., there would be no...</i>	1	A2
	likely to	<i>...that the program is not likely to be successful when applied to syntax.</i>	2	A1
	must (modal)	<i>...there must be a rich innate language-specific universal grammar...</i>	1	A4
		<i>There must be at least one level of...</i>	4	A4
<i>There must be some point in the mapping...</i>		4	A4	
observe	<i>We can now also observe...</i>	1	A3	
reasonable	<i>It is both reasonable and economical to...</i>	1	A5	
	<i>This is eminently reasonable...</i>	3	A5	

EVIDENTIALITY STRATEGIES AS DISTINGUISHING MARKERS ...

seem	<i>...and the way to do this, it seemed, was to ...</i>	1	A3
	<i>There seems to be...</i>	1	A4
	<i>These ways of phrasing the question seem nearly indistinguishable.</i>	3	A4
	<i>...Fodor seems to be coming down on...</i>	3	A4
	<i>...sentences like (1) don't seem to bear on anything normally considered syntactic ...</i>	4	A4
	<i>It seems perfectly natural to claim that...</i>	1	A5
	<i>...new approach after new approach in phonology seems to rally most of the field behind it...</i>	2	A1
	<i>...it seems accurate to use the expression...</i>	2	A1
	<i>FOT seems incompatible with...</i>	2	A1
	<i>It might seem surprising that...</i>	2	A1
	<i>It seems to me beyond question that...</i>	1	A2
	<i>...they seem quaint...</i>	3	A2
sound	<i>Option B sounds like something...</i>	4	A4
	<i>Option C may sound a little mysterious ...</i>	4	A4
thus, therefore, thereby, then, so	<i>...indeed, thereby constitutes the main evidence for grouping...</i>	2	A3
	<i>Thus, this...expresses the path...</i>	2	A3
	<i>Overall, FO, thus emerges as...</i>	4	A3
	<i>Thus at one extreme we find...</i>	1	A4
	<i>The conclusion, then, is to...</i>	1	A4
	<i>NPN thus provides strong arguments...</i>	2	A4
	<i>So, the issue can be stated more precisely...</i>	4	A4
	<i>The statue rule is therefore a principle...</i>	4	A4
	<i>Categorization is then a matter of...</i>	3	A5
	<i>The possibilities... are therefore double...</i>	2	A1
	<i>Not surprisingly, then, recent years have seen...</i>	4	A1
	<i>Thus, the issue of whether global rules...</i>	1	A2
<i>...thus less likely to serve as a prime.</i>	3	A2	

III–General knowledge evidentiality – lexico-grammatical expressions and examples

	Lexico-grammatical item	Example	Sample	Author
g e n e r a l k n o w l e d g e	fact	<i>The fact that this second alternative is...</i>	2	A3
		<i>The fact that language and music are...</i>	1	A4
		<i>The fact that they are conventionalized...</i>	1	A5
		<i>The fact that speakers mentally represent...</i>	1	A1
		<i>The fact that grammatical change is...</i>	1	A1
		<i>The fact that speakers can make...</i>	1	A1
	It is agreed that...	<i>It is widely agreed that aspects of linguistic structure can be...</i>	1	A5
		<i>It is further agreed that these functions have...</i>	1	A5
	It is well-known	<i>It is also well-known that... Check and Slovak linguists were...</i>	4	A1
	true	<i>It is not only true but also a truism that...</i>	4	A1
	view	<i>In the orthodox view, basic grammatical categories are defined...</i>	3	A5
		<i>Traditionally dominant has been the view that a category is defined...</i>	3	A5
		<i>...given the predominant view that the...</i>	2	A1
	was/were seen [in the context of a theory]	<i>Concepts were seen as characterized by...</i>	3	A2
		<i>The mind was seen as manipulating abstract symbols without...</i>	3	A2
		<i>Language was seen through the Chomskian metaphor...</i>	3	A2
<i>In classical theories of language, metaphor was seen as a matter of language not thought.</i>		4	A2	