Renata Pípalová

Charles University in Prague, Department of English Language and Literature

Interweaving citations in academic discourse by (non)native (non)professionals

This paper studies manifest intertextuality, namely (free) direct speech in academic discourse. To that end, a corpus was assembled consisting of four samples of professional academic prose written by native speakers, four samples of professional academic prose written by non-native Czech linguists and four samples written by non-native (Czech) undergraduates. The research has two specific foci: initially, attention is given to the discourse parameters of academic citing (i.e., who is quoting whom, from where, what, how frequently, etc.), and later, the research aims at a range of framing structures, interweaving the citations in the current text. The paper offers a comprehensive analysis of reporting frames and related structures introducing citations in academic writing, examining their range, positions, the subjects featured, the word order, type of verbs, etc. The findings of this paper, comparing tendencies in native and non-native samples on the one hand, and in professional and novice discourse on the other, may be of use in academic writing courses at universities.

Keywords: academic discourse, reporting frame, citation, EFL

1   Introduction

In line with most research in English for academic purposes (EAP) (see, e.g., Swales 1990; Bhatia 1993; Hyland 2000, 2006; etc.), this paper investigates written academic discourse exclusively. More specifically, it studies its manifest intertextuality as it is shown in reporting. Reporting (direct speech, reported speech, citing, quoting, represented discourse, etc.; see, e.g., Leech & Short 1981; Fludernik 1993; Lipson 2006; Hoffmannová 2008; Keizer 2009; Brendel, Meibauer & Steinbach 2011; Johansen 2011; Pípalová 2012) constitutes one of the defining characteristics of academic discourse, and serves a number of functions. To name but a few, it displays the writer's concern for interactions with an audience, has a persuasive function, appeals to the community's shared knowledge in order to build firm ground for the writer's line of argumentation, creates a rhetorical niche in research, acknowledges a debt of precedent, invokes the respective epistemological and literacy traditions, facilitates the writer-reader interaction, cements cooperation and disciplinary peer relationships, helps create a dialogic space for the introduction, negotiation and acceptance of claims, etc. (see, e.g., Hyland 2000).

This paper focuses solely on manifest intertextuality where the cited text is explicitly present in the sample under analysis and must also be clearly marked graphically, even though it is frequently inseparable from constitutive intertextuality (i.e., configuration of discourse conventions that go into its production, see Fairclough, 1992: 104 cited in Hyland 2000: 21). According to Hyland (2000: 21),

> explicit reference to prior literature is a substantial indication of a text's dependence on contextual knowledge and thus a vital piece in the collaborative construction of new knowledge between writers and readers. The embedding of arguments in networks of reference not only suggests an appropriate disciplinary orientation, but reminds us that statements are invariably a response to previous statements and are themselves available for further statements of others.

In a single English-language academic study, the author may cite themselves or others, they may refer to one author or to many, once or repeatedly, using a single source or numerous sources, in the original or in translations, they may cite their contemporaries as well as academics from previous generations, scholars coming from diverse language and cultural backgrounds, from various schools and epistemological traditions, using varied codes (e.g., ENL, EFL, or other languages), etc. Thus, academic English is constantly exposed to new incentives and enrichments from a web of local discourses.

This paper explores manifest intertextuality in linguistic discourse. Actually, linguistics is a rather special field, as authors may cite secondary, as well as primary sources, and from the secondary sources they may quote examples rather than metalinguistic

passages. Depending on the topic, in the absence of appropriate research, authors may occasionally cite diverse non-academic sources, e.g., the popular press. Hence linguistic discourse is a par excellence embodiment of what Bakhtin (1980) calls heteroglossia.

This paper aims to explore linguistic discourses in view of two non-regional variables: code (native vs. non-native English) and experience (discourse by professionals and novices). The choice of these foci is meant to reflect some of the significant characteristics of the vast, diffuse and constantly changing academic community. Members of the academia, who as a rule do not know all the members personally (imagined communities), usually establish weak ties. As Mauranen, Hynninen and Ranta (2010: 11), following Milroy, note, "networks of weak social ties tend to favour linguistic innovation and dialect levelling, in contrast to networks where strong ties dominate." It is due to academic mobility, scholars' participation in global networks of research, etc., that at present English serves as a Lingua Franca (ELF) for world-wide academic interactions and, in quantitative terms, the non-native academic discourse naturally prevails over its native counterpart. Indeed, academia has been dominated by non-native performance, which may explain the striking attention recently given to EFL (e.g., Mauranen *et al.* 2010; Bjorkman 2011; Ferenčík 2012; Povolná and Navrátilová-Dontcheva 2012). As a rule, ELF arises in multilingual communities. According to Mauranen *et al.* (2010: 11), "this is very clear in universities: the matrix culture is frequently one where English has no major status, and speakers invariably possess other language resources." It should be recalled that citation is an integral part of the intended reader's expectations and is indispensable in particular sections of academic discourse. Therefore, familiarity with its techniques has become a firm element of novice socialization into the academic discourse community and is given thorough attention in diverse style manuals.

2 Data

The research corpus was composed of twelve relatively recent academic samples (all published within the last fifteen years). There were four monographs written by native speakers of British English who were also professional linguists (English Professional Subcorpus, hereinafter **EPS**); there were four monographs by non-native speakers, i.e., Czech professional Anglicists (Czech Professional Subcorpus, hereinafter **CPS**); last of all, there were also four final projects produced by non-native undergraduates (Czech students, Anglicists) in their final years of study (Czech Student Subcorpus, hereinafter **CSS**). Hence the non-native subcorpus comprised four samples by renowned linguists of different generations, and four samples by students, novices to the academic community, the diploma thesis being their very first attempt at an academic study. A

complete overview of the subcorpora samples is provided at the end of this study. Since the sources in the corpus differed in the frequency of the phenomenon in question, to ensure comparability of data, I extracted the first 25 specimens from each. Thus the corpus was composed of 300 excerpts altogether, each of the three subcorpora comprising 100 specimens. Naturally, the size of the corpus cannot but provide preliminary insight, which should be tested further.

As to the criteria, to count as a specimen, the citation itself had to be at least one sentence in length, syntactically self-contained and sufficient, whether or not the first letter was capitalized. Hence the quoted passage had to be syntactically complete, featuring a capital or lower-case letter at the beginning and a terminating stop in the end, or possibly several stops at the beginning and/or in the end to suggest deliberate ellipsis. The boundary was marked graphically (by single or double quotes, by distinct fonts, or else by a set-off paragraph). Last but not least, the reporting clause or a similar immediately preceding structure clearly suggested the source, creating an explicit intertextuality link to another discourse. Occasionally, though, such a structure was missing and was compensated for only by the bracketed cross-reference to the source, or to the footnotes or endnotes giving such information. That said, however, the source also had to be listed in the bibliography or references section of the citing sample. Hence my specimens complied with what Leech and Short (1981) call Direct Speech or Free Direct Speech, disregarding other structures suggesting intertextuality (e.g., Mixed Citations or Indirect Speech passages).

3   Findings

The research was done in two steps. Initially, attention was given to some of the prominent discourse parameters of citations, such as who is citing whom, what is cited, from which source or genre, which medium and channel are employed, how frequently, etc. (3.1). Later on, attention was shifted away from the citation to the stretch introducing it, as the paper examines the ways the authors employ to integrate citations in their texts (3.2, 3.3).

3.1   Conspicuous discourse parameters of citations

In their quotes drawn from secondary literature, the professionals cited their peers, while the novices cited renowned members of academia. It follows that the social distance between the undergraduates and cited experts was rather maximal. Surprisingly, self-citations were not detected in the corpus at all. Since most authors cited directly from

sources, mediated citation turned out to be very rare. Nevertheless, it was detected in all the subcorpora, with somewhat higher incidence in the novices' data.

The number of authors quoted per sample or per subcorpus varied, and the same holds for the number of cited sources. There were samples with a striking cited-author turnover, and samples focusing on some authors and/or titles. On average, each author appeared twice per sample and was represented by one source. However, striking differences surfaced. Only to mention two extremes, one sample in the CSS quoted five authors and sources. Conversely, in the EPS, a sample referred to nineteen distinct authors and sources.

Nearly all the corpus citations were formulated in English, the majority coming from speakers of ENL (79%). Only a small proportion was phrased by speakers of EFL (20%). These obtained chiefly in the non-native subcorpora, and over a half of them referred to Czech producers. This presumably suggests the affiliation with the local community of practice and respect to its traditions. In contrast, in the EPS, non-native speakers were cited rarely (8%), and in the official translations only. There were also some exceptions (1%) worded in Czech, with no English translation, all drawn from the CSS.

Nearly all the specimens (95.34%) came from printed sources, and the remaining 4.66% from electronic ones, and from the CSS only. Tellingly, a decisive share of the instances (78.66%) were gathered from secondary literature, with a fifth (21.34%) coming from primary sources. A clear majority of the secondary sources (64.67%) involved metalinguistic comments, the rest (13.39%) embodying examples. Admittedly, despite frequency differences in individual samples, these figures are influenced by the fact that the first 25 instances were drawn, which prioritizes the conventional theoretical sections of monographs. Interestingly, the highest proportions of primary source citations were detected in the EPS, and the greatest share of secondary literature examples came from the CSS.

As a rule, the primary sources and examples from secondary literature came from non-academic fields, ranging from religious, via journalistic all the way to fiction discourse. There were quotations from a variety of genres, such as poetry, drama, face-to-face conversation, prose, etc. The citations featured examples of spoken, as well as written language, and were originally provided by graphic, phonic or electronic channels. Understandably, secondary sources were dominated by academic linguistic discourse from a range of genres, including monographs, grammars, dictionaries, journal articles, etc.

Citation frequency was also found to vary immensely. It proved to be the densest in the CSS, and the least frequent in the CPS. Indeed, the 100 specimens were drawn from approximately 1,400 words of the CSS, from around 1,700 words of the EPS, and

from nearly 2,300 words of the CPS. Regarding length, typical direct speech featured 3–4 lines, although the shortest covered a single line, and the longest spanned fifteen lines. Students generally preferred longer citations (4.11 on average), whereas professionals tended to feature shorter ones (EPS 3.93 lines, and CPS 3.38 lines). This may also have to do with the courage to highlight certain points in citations through necessary abridgments.

Actually, professionals often dared to modify the citations (i.e., by deleting a passage or by inserting an explanatory note) when necessary, of course providing that such modifications were explicitly marked. Moreover, they frequently suggested the hierarchy between their texts and the reported discourse even graphically, giving the quotes in smaller fonts, italics, etc., presumably following the style sheets. Indeed, modification of citations arose more frequently in professional writing, as 35% of quotes were modified in the EPS, 19% of instances in the CPS, and only 14% of citations in the CSS.

As for modification tendencies, all the subcorpora unambiguously favored abridgements, although the non-native speakers nearly exclusively so. Apparently, non-natives in general, and the novices in particular, showed greater respect to authority and found it more difficult to modify the formulations. Thus the more pro-active modifications (for example, inserting an explanatory note or providing a paraphrase) featured chiefly in the native subcorpus. Such modifications were presumably conceived of as cooperative acts enhancing coherence and saving some processing effort.

Considering the graphic aspect, many authors found it useful to distinguish graphically between metadiscourse citations on the one hand, and the primary source quotes together with illustrative examples from secondary sources on the other, employing italics in the latter cases. Some authors even set off all the citations in italics, irrespective of the source. As for the punctuation marks, inverted commas featured frequently, but mainly with in-text citations, whereas in longer quotes, authors frequently preferred other signaling ways. They mostly changed fonts (to italics or smaller ones), narrowed paragraphs and skipped a line instead. English writing favored single inverted commas, while the Czech data (the CPS and the CSS) preferred double inverted commas, which presumably follows from the conventions established in the Czech community of practice.

Generally speaking, novices proved to be less careful editors, occasionally resorting to the Czech signaling conventions instead (introducing citations without raising the inverted commas above lines at their onset). At times, their use of punctuation devices appeared more confusing, with frequent stops placed instead of expected colons. Conversely, the professional data turned out to be more user-friendly, prioritizing colons in more than three quarters of instances. In professional writing, native and

non-native alike, resulting from more convincing editing and/or implying more hands, inconsistencies were nearly missing.

3.2　Source referencing

To avoid plagiarism, the source of citation has to be provided explicitly. By the established convention, it has to be identified (at least) through the surname of the author, year of publication, and page number, although in addition to the foregoing, other details may be given (e.g., first name of the author, publishing house, title of the work). Referencing information may be included in the reporting clause or a similar structure, or it may take a non-sentential, bracketed form, or else, authors may choose to combine the two.

Whenever the framing construction was actually coupled with its bracketed surrogate, the latter presumably served also surveyability purposes. Although in such cases information redundancies were frequently detected, bracketed reference was prone to provide fuller information and served also as a terminating device as the following example illustrates:

(1)　Hull writes, 'People reject models in conceptual change out of hand because they have a simplistic understanding of biological evolution' (Hull 1988:402) (Croft, 12)

In fact the data in the RC or its bracketed surrogate (i.e., a frame of a kind), cross-refer not only to the source itself (intertextuality), but along with it, to the Works Cited section in the citing text, and sometimes also to its footnotes/endnotes (intratextual links). To be sure, in a sample, the source may be quoted more than once, these citations forming cohesive chains of diverse lengths. Hence cited passages enter a complex network of links.

As to the position of the obligatory items, all the data could be placed before the citation (anteposition, typically enclosed in a reporting structure, 16.33%), inside the citation (medial position, 0.33%), or after the citation (postposition, 39.67%). Occasionally, some data came before and some following the quote (ante-postposition, 43.67%), involving potential information redundancies. In the examples, relevant passages are underlined:

(2)　Ante-position: As Isard (1975.287) remarks: "Communications ....(abridged by RP)...." (Povolná, 16)

(3)　Ante-postposition without information overlap: Trask defines presupposition as follows: 'A proportion..(abridged by RP)...sensible' (1997:175). (Wilson, 39)

> (4) Ante-postposition with information overlap: <u>Keller</u> describes this desideratum as the principle of methodological individualism: 'the explanation... .(abridged by RP)... or collectives.' (<u>Keller 1990/1994:121</u>). (Croft, 4)

> (5) Postposition: Others do not recognize the distinction, leading to misunderstandings such as that in the following passage: How can one understand .....(abridged by RP)... 'actor'. (<u>Lass 1997: 339</u>) (Croft, 5)

Since in the favored patterns, i.e., the ante-postposition and the postposition, the fuller information in brackets came after the quotes, there seems to be an overwhelming tendency to locate fuller information finally, presumably also due to the end-weight and end-focus principles (see, e.g., Greenbaum & Quirk 1990). Regarding the ante-postposition, a lesser proportion of the specimens avoided information overlap (13.67%), while a more sizable group did involve redundancies (30%). Scrutiny shows that the professionals preferred to do justice to language economy, whereas the students did not. This may follow from the novices' insecurity as to adequate compliance with the citing conventions.

3.3   Framing structures

3.3.1 Reporting clauses

The reporting clauses (RC), the simplest introducer of quotations, proved to be the dominant verbal reporting form, constituting a lesser half of all the corpus instances (43%). Unlike their bracketed surrogate, which came mostly in final positions, RC were regularly placed initially. They were more abundant in the CSS (49%) compared to both its professional counterparts – the EPS (41%) and the CPS (39%). Apparently, the professionals preferred formal diversity.

In all the subcorpora the leading pattern (nearly 65% of RC), proved to be SV(O), i.e., subject & transitive verb, with direct object constituted by the citation itself:

> (6) In the same period, a great humanist Abraham Lincoln, the 16th president of the United States, famous for leading his country through the Civil War and for ending slavery for good, wrote in 1858: (Ženíšek, 29)

The second most productive pattern (29%) turned out to be in fact syntactically self-contained, i.e., SVO. In such instances, the relationship between the citation and the reporting frame corresponded to apposition constituted between the citation and one of the clause elements in the frame (i.e., O, S, or the optional modifiers in question):

(7)   In a paper on the status of 'non-sentences', Stainton makes a convincing defense of the thesis that 'Ordinary words... ' (abridged by RP) (Wilson, 16)

This pattern was detected chiefly in the non-native data. All the other patterns proved to be rather marginal (6% altogether).

Since all the frames in the corpus displayed direct word order, there was no subject-verb inversion detected. Almost all the frames were positioned initially. The only exception placed medially turned out to be interesting also in that the citation itself employed Czech:

(8)   "Vlastní jména," Dušková (1988) says, "přecházejí k apelativům ... (abridged by RP) (Huserek 13)

A clear majority of subjects in frames were animate ones. The most numerous groups of subject realizations included antroponyms (44.18%), personal pronouns (23.26%), and inanimate nouns (11.63%), although other less frequent forms were also discovered (for example, common personal nouns, demonstratives, indefinite pronouns, 20.93% altogether). As to the inanimate subjects, they usually combined with passive verb forms. The following instances illustrate a range of subjects in frames:

(9)   Sinclair and Coulthard (1975) say: (Kovačová, 10)

(10) He adds, for instance, that (Croft, 11)

(11) Examples such as these are fairly common in the corpus: (Hyland, 34)

Surprisingly, lexical verbs displayed a pronounced diversity, as the 129 RC employed in all 55 verbs. It follows that as a rule, a verb featured approximately twice (2.35 tokens per verb). The complete list of verbs is as follows: *add, adduce, agree, analyse, argue, ask, boast, capture, characterize, claim, comment, compare, complain, consider, delete, define, describe, discuss, distinguish, elaborate, emphasize, excerpt, exemplify, explain, find, formulate, give, hit at, illustrate, indicate, maintain, make a defense, mention, note, point out, present, propose, quote, read, recognize, reflect, reply, respond, say, state, stress, summarize, sum up, support, take, take the view, write, urge, use, utter.*

The most frequent verb proved to be say, detected in 11%, largely non-native frames, followed by *claim, write, maintain, and argue*. The number of tokens per verb was minimal in the English data (1.7), compared to both the Czech counterparts (1.81 in the CSS and 2.0 in the CPS). Apparently, the non-native subcorpora exhibited some degree of stereotypicality in frames. However, the reasons may differ, as in the novices, it may follow from their ongoing socialization process into academia, whereas in the professionals,

it may be also due to a striking use of more demanding referencing structures, which alone is likely to relegate the reporting frames to cliché status. Moreover, the EPS favored verbs associated with formal, written discourse, whereas the non-native subcorpora preferred less formal, spoken items. Indeed, apart from the dicendi type, the cogitandi verbs were abundant, especially in the EPS (*define, recognize, capture, reflect, exemplify*, etc.). The native data were also dominated by verbs connoting literacy (*recognize, write, characterize,* etc.), while the Czech subcorpora featured some orality items instead (*say, give, use,* etc.). Inside the CPS, however, a disproportion arose, which roughly paralleled generation membership and/or idiolects, since younger linguists were more inclined to use verbs connoting informal, spoken language, presumably to reflect such trends in current academic discourse.

Regarding the verb forms, present simple dominated in all corpora (over 75% of the frames). It proved to be nearly the only verb form in the CPS subcorpus (89%), its incidence was mean in the native subcorpus (76%), but it was rather underrepresented in the novices' data (63%). Thus, professional samples displayed greater morphological uniformity, whereas the novices' data showed noticeable formal diversity. The remarkable rate of past simple in CSS may well be due to the transfer from the mother tongue.

Around a half of the frames were modified by optional clause elements. Wherever the modification was richer and more conspicuous, the frames came close to similar framing structures to be discussed herein below. Approximately a half of all frames were devoid of modification, whereas a half did involve optional constituents. Interestingly, the subcorpora prioritized distinct types of modifications. Whereas in the professional subcorpora, it was usually an optional adjunct that was integrated in frames, the novices favored optional attributes and appositions (6). Unlike the native professionals, who kept multiple modification in frames to the minimum, the novices were by no means reluctant to employ it in 22.45% of instances, with the CPS findings falling somewhere between the two. On the whole, modifiers in frames were used by far the most sparingly in the EPS. This may well be connected with the more straightforward, direct and fact oriented English writing style. Conversely, as suggested e.g., by Čmejrková (1996) or Kaplan (2000), the Czech academic tradition seems more verbose, convoluted and baroque.

Typically, a reporting frame and the quote made a separate sentence. It was rather rarely that the frame became part of a more complex syntactic structure, featuring as a clause in a compound sentence (12), or outright as a finite (13) or non-finite (14) subordinate clause:

(12) Mufwene compares languages to species and the factors that determine a language's survival or extinction as ecological factors, but states: (Croft, 11)

(13) The **effect of co-text on the interpretation of what follows** has been studied, for instance, by Brown and Yule (1983.59), who maintain that (Povolná, 17, bold in original)

(14) Halliday advocates the presence of structure in spoken language stating that (Urbanová, 12)

There were six such complex instances in the CSS, eight cases in the EPS, but a striking seventeen specimens in the CPS. This phenomenon apparently indexes a more demanding style, which strives to hierarchize, integrate and condense its content.

3.3.2 Stance adverbials

A close relative of reporting clauses is stance adverbials (SA), see, e.g., Biber, Johansson, Leech, Conrad and Finegan (1999). With the RC, they share some noteworthy features, e.g., they typically go before the citation, they may exhibit a comma or colon at the end, they may include a referencing bracket, etc. Moreover, like the frames, they tend to display dicendi or cogitandi verbs, usually in present simple active forms, although some may be verbless:

(15) As Vygotsky points out: (Lantolf, 25)

(16) As he puts it, (Adam, 33)

(17) According to Schiffrin (1994: 5), (Kovačová, 6)

As to the incidence of the SA, they constituted 11% of all the specimens. They were found to be fairly prolific in the English data (18%), while noticeably marginalized in Czech writing. More specifically, in the CPS their frequency dropped down to a half (9%), and in the CSS they turned out to be rather rare (6%). Regarding the forms, the *as*-construction convincingly outnumbered the *according to*-structures in all three subcorpora.

3.3.3 Other framing structures

In addition, the corpus included a considerable share of other framing structures. I labeled them as follows: Anticipatory Indirect Speech introducing citations (AIS), which usually involve nominal content clauses and may comprise some of the authentic expressions, i.e., Mixed Citations (18); Anticipatory Reports of Speech Acts (ARSA), which inform about the implementation of a speech act and label it (19), and Content Anticipators (CA), i.e., paraphrases, metadiscourse comments, anticipatory evaluations of the coming content, general claims to be illustrated, etc. (20). Admittedly, at times it could be difficult to neatly distinguish between such anticipatory structures, as they form a relatively continuous cline.

(18) Lyons (1995.229) admits that his (and others') attitude towards the notion of **grammaticality** and **semantic well-formedness** has undergone substantial changes: (Urbanová, 19)

(19) Hull quotes an organism selectionist, Mayr, and then a gene selectionist, Dawkins: (Croft, 22)

(20) The following quote nicely captures Vygotsky's thinking on the matter of symbolic artifact: (Lantolf, 26)

These structures share at least the following features: They are all syntactically self-contained, but semantically richer than RC. They represent sentence introducers to citations, meant to ensure smoother integration of the quote in the current discourse and to enhance its coherence. Typically, they precede the citations and are terminated by colons. They tend to display cataphoric means and incorporate explicit reference to the source text. The verbs they involve resemble the sets employed in RC and SA. Many are of dicendi and cogitandi types, preferably formal and learned ones, frequently of foreign origin (e.g., *evaluate, demonstrate, justify, advocate, recollect, exemplify*), employed usually in present simple forms. These structures being syntactically self-contained and semantically richer, compared to the RC, enables the authors, among others, to express in them their high point, to prepare the reader for the coming content by means of its anticipatory paraphrase, to interpret or evaluate it in advance, or else they make it possible for the reader to rely on the reformulation and skip a lengthy citation altogether. Although the structures are syntactically sufficient and complete, semantically, however, they are not, as they call for listing, particularization, exemplification, etc. This fact is suggested, among others, by the typical final colons.

As to the findings, all the corpora concurred in giving the greatest priority to the CA, although the professional data turned out to be far more convincing in this respect. Indeed, whereas in the EPS the CA corresponded to 23% of data and in CPS they complied with 26% of specimens, they were clearly underrepresented in the CSS, covering solely 11% of all. The results for the other anticipators were far less significant and contrastive, as the AIS constituted 4% and ARSA 6% of all. Interestingly, however, there were also some combined forms detected. Typically, a CA coupled with a RC or SA, although other combinations were by no means precluded. Such complexity surfaced in 8% of the EPS instances, 11% of the CPS examples and in 6% of the CSS data. Example (21) illustrates AIS and RC, example (22) shows ARSA and RC, and example (23) displays CA and RC:

(21) He argues that the traditional designations of both the articles – "definite" and "indefinite" – are unsuitable and gives the following explanation: (Huserek, 7)

(22) He elaborates on the problems associated with this definition and notes: (Wilson, 17)

(23) Halliday advocates the structure in spoken language stating that (Urbanová, 12)

Apparently the professionals strived to exercise greater control over the readers' processing of quotations, they intended to enhance coherence of the text, to ensure nearly a seamless intertextuality link and they also strived to foreground content matters at the expense of more formal cross-referencing data, providing the reader with more profound background knowledge. Conversely, the non-native novices seemed rather reluctant to analyze and interpret the citations. Thus they chose to exercise far less control, frequently leaving the responsibility for the reception upon the recipients.

Paradoxically enough, over 7% of the specimens displayed only implicit frames (IF), which in fact gave rise to free direct speech instances. This was the case when the quotation followed a (sub)title, or when the frame was ellipted owing to coordination or branching, and when the citation was simply juxtaposed side by side to another citation:

(24) Hull writes, 'If ever anyone thought…..(abridged by RP)…metaphor to rest' (Hull 1988:218; see Hull 1988:442; Mayr 1982:794-807; Dawkins 1982b:85-6 for more details), and 'in both biological and conceptual evolution,..(abridged by RP)…There are no unit genes or unit ideas. ' (Hull 1988:449) (Croft, 28)

However, the subcorpora differed quite radically in this respect. Indeed, the IF were found to be scarce in the EPS (2%) and in the CPS (4%), but in the CSS, they constituted the second most prolific pattern (17%). In both the professional subcorpora, IF stem chiefly from integrating citations in more complex syntactic structures (e.g., compound citations), while in the CSS their use may follow from the endeavor to eliminate an excessive share of stereotypical RC and/or from the intention to ensure graphic surveyability.

4   Discussion of findings

Research showed that in the corpus, a typical intertextuality link referred directly to an original secondary literature source, published recently, by a native speaker of British English, and made available in a hard copy. The length of a citation fell usually between 3–4 lines, the quote being marked graphically. Prototypically, in a single corpus sample, each author was cited twice, using one title. Given the composition of the corpus, peer-citation clearly outnumbered non-peer quotes. The mere fact that not a single professional cited a novice, whereas novices cited experts and never their peers, suggests hierarchy, and an imbalance of knowledge and authority. Furthermore,

there was a pronounced tendency towards congruence between the cited and citing samples in terms of the language, medium and channel employed, but incongruence of genre. Since each corpus sample of 25 instances comprised on average reference to 12.6 other authors, and none did to themselves, manifest intertextuality makes academic discourse a perfect embodiment of what Bakhtin (1980) calls heteroglossia. The field selected being linguistics, authors cited secondary, as well as primary sources. Moreover, from the secondary sources, they occasionally employed only illustrative examples. The citation frequency, and the cited-author and cited-source turnovers in samples varied immensely, which reflects a number of factors, including the communicative intention of the author, the availability of previous research on the topic, the author's familiarity with such research, the focus put on analyzing authentic examples or overviewing or assessing previous research, preference of other forms (e.g., mixed citations), and position in the monograph.

Since the corpus explored was relatively restricted, naturally, the conclusions drawn may be only tentative. Nevertheless, the corpus was large enough to show 1) that there is a striking tendency to position all the verbal framing structures initially, which presumably stems from the thematic function of frames and which in turn unambiguously assigns the citations their rhematic status; and 2) that a wide variety of structures may be employed to introduce intertextuality links, which include, apart from the well-known RC, also the SA and communicatively much richer forms, some of which themselves have traditionally been known to refer to other sources, such as AIS, and (an academic counterpart of) NRSA. In addition to these, this paper introduces content anticipators. At the minimum, however, the frame may be compensated for by the bracketed surrogate.

In fact, the framing structures may be arranged along tentative syntactic and semantic RC (with a rather indistinct syntactic status in diverse treatments, varying from main clause standing to attitudinal disjuncts/comment clauses, see, e.g., Dušková *et al.* 1994; Quirk, Greenbaum, Leech & Svartvik 1985; Biber *et al.* 1999; Huddleston & Pullum 2002; Keizer 2009), via syntactically sufficient RC featuring objects and potentially embracing ample optional modification, all the way to a clear content paraphrase, syntactically self-contained and semantically much richer (taking the form of ARSA, AIS or CA). In such cases the reader can in fact posit a loose appositive relation between such introductory paraphrase and the citation itself, be it the relationship of full or partial equivalence, or others.

Thus, the line separating the semantically basic and syntactically insufficient examples from the self-contained and elaborated ones is somewhat fuzzy and continuous. In addition to providing solely the conventional referencing data, these structures may also particularize some unusual elements of the reported situation and

anticipate the content of the coming citation. It was chiefly in professional discourse that such self-contained structures featuring the direct object proved to be numerous, presumably due to the endeavor of the authors to ensure a particular reception, guiding the reader, constraining room for interpretation, reinforcing meaning, and facilitating understanding by preparing the recipient for the citation in simple words, etc.

To summarize, the quantitative results of this research, the RC proved to be the leading device (43% of all). The second most significant framing structure turned out to be CA, usually coupled with bracketed frames (20%), followed by SA (11%). Table 1 provides an overview of all the corpus findings. From the table it follows that the EPS came closest to average values. The non-native corpora took, in fact, two extreme positions. Whereas the CPS prioritized more demanding, semantically and syntactically sufficient frames, the CSS did the opposite, favoring the syntactically indeterminate, semantically basic, or even implicit ones.

TABLE 1.   Corpus findings in the EPS, CPS, CSS regarding Reporting Clauses (RC), Stance Adverbials (SA), Anticipatory Indirect Speech (AIS), Anticipatory Report of Speech Act (ARSA), Content Anticipators (CA), Implicit Frames (IF), Combined Forms (COMB).

	EPS	CPS	CSS	abs. TOT %	TOT
RC	41%	39%	49%	129	43%
SA	18%	9%	6%	33	11%
AIS	3%	3%	6%	12	4%
ARSA	5%	8%	5%	18	6%
CA	23%	26%	11%	60	20%
IF	2%	4%	17%	23	7.7%
COMB	8%	11%	6%	25	8.3%
TOT	100%	100%	100%	300	100%

Thus, professional writing seemed rather focused and more likely to be perceived as coherent, for even the graphic means were serviceable to it. Professional monographs displayed a noticeable tendency to duplicate and reinforce the message, by way of employing a considerable rate of anticipatory structures, which also resulted in deliberate disambiguation of quotes and testified to the author's cooperative guidance. That said, anticipators of all kinds constrain room for free interpretation, facilitate a particular processing of quotes and enhance academic explicitness and accuracy. Apparently, professionals preferred structures where more information could be packed in.

Naturally, undergraduates being novices to academia, their citations necessarily acknowledged hierarchy, respect and recognition. They may have found it demanding

to paraphrase, interpret or modify (e.g., abridge) the quotations. Presumably, they may not have been fully aware of orality – literacy norms and conventions. The forms they preferred resulted from a number of factors (e.g., age, experience, non-native status, degree of socialization, less clear focus, interpretative caution, avoidance of interpretative imposition and constraints). Also, they have grown in, and are part of, the Czech academic community of practice, having been schooled primarily by Czech tutors.

In this connection, it may not be without interest to look at the peculiarities of this tradition more closely. For example, Čmejrková (1996: 144) following Clyne, maintains that the Czech academic register has many features in common with German scholarly discourse:

> Texts written by Germans (...) are less designed to be easy to read. Their emphasis is on providing readers with knowledge, theory, and stimulus to thought. In English speaking countries, most of the onus falls on writers to make their texts readable. (...) In German-speaking countries, it is the readers who have to make an extra effort so that they can understand the texts. This presupposition that it is the reader's responsibility to understand rather than the writer's responsibility to write it understandably also seems to be deeply rooted in the Czech stylistic tradition.

To explain the tendency, Čmejrková (1996: 145) offers several considerations, including the following: "This feature has various sociolinguistic motivations in Czech: Czech readers have been trained to read even between the lines and to infer the sense that is text immanent".

Indeed, the Central European academic community of practice has been marked by relatively non-interactive, intellectually demanding and convoluted prose. Among others, this may well explain that the CPS authors favored CA, which appear to enhance the content compactness far more conspicuously, even more than did the native speakers.

There seems to be very little that the two non-native subcorpora share, presumably with two exceptions. Firstly, they exhibit a more restricted range of forms. For example, both clearly favor RC at the expense of SA. Secondly, they seem to suggest, although in different degrees, that the Czech culture displays a greater tolerance to implicitness, having been for so long, in the past, trained to read between the lines. Naturally, the degree of proficiency, professional erudition, together with the regular practice of native-speaker editing prior to publications may have smoothed away numerous non-native features from the Czech professional discourse. Nevertheless, some formulation complexity can be attributed, apart from cultural specificities, also to non-native status. As Mauranen *et al*. (2010) show, EFL affects even metadiscourse and rephrasing peculiarities, stemming from the endeavor to enhance clarity and explicitness. Arguably,

such tendencies may be traced even in the novices' data, chiefly in the rate of optional modification in their frames.

5   Conclusion

To conclude, the actual choice of the reporting structure stems from an interplay of numerous factors, including the code (ENL or EFL), idiolect, generation, academic culture and community of practice, degree of socialization into academia, length of citation, its formulation complexity, and editing (more hands implicated), etc. Moreover, the framing structure chiefly follows from the producer's specific intention: their attitude to the recipient and to the citation. Hence it reflects their cooperation with the reader, their control over the reception of the quote and the degree of imposition, or their respect to the reader's interpretative autonomy, their attitude to the role and significance of citation, their need for its assessment, for taking a viewpoint, emphasizing a crucial aspect in order to integrate the quote smoothly into their prose, etc. Among others, the type of structure also shows how much the author thinks it indispensable for the reader to actually process the citation or whether they offer a comfortable paraphrase instead, which gives the reader an option to skip the quote and rely on the reformulation only.

Naturally, the results of the present study are only preliminary. More valuable insights could be gained by examining a larger corpus, and including student data from outside Czech universities. Nevertheless, since the results gained from the non-native professional subcorpus came closer to the native tendencies than did the data drawn from the non-native novices' subcorpus in a variety of respects, it only seems natural to conclude that the socialization process into the academic community is a long one and that should they want to, students have yet a way to go to come to terms with some of the strategies and conventions in order to be well accepted by the international academic community.

To suggest some applicability of the aforementioned results in courses of academic writing, teachers should cultivate the students' awareness of distinct communities of practice and related literacy and orality norms and conventions. Moreover, they should draw the students' attention to significant transfer risks, etc. To enhance the reader's perception of coherence, students ought to be encouraged to hierarchize their content more conspicuously, to reduce the number of implicit frames and to employ a range of reporting structures. Since the novices are being socialized into the academic community, the guidebooks and manuals may prove a valuable source of information and thus they should be systematically made to use them. Last but not least, students

should be trained to be more consistent and careful editors, paying sufficient attention even to the graphic aspect.

References

Primary sources

English Professional Subcorpus (EPS) (Native linguistic monographs):
Croft, W. 2000. *Explaining language change. An evolutionary approach.* Harlow: Pearson
    Education.
Hyland, K. 2000. *Disciplinary discourses. Social interactions in academic writing.* Harlow: Pearson
    Education.
Lantolf, J. P. & S. L. Thorne 2006. *Sociocultural theory and the genesis of second language
    development.* Oxford: Oxford University Press.
Wilson, P. 2000. *Mind the gap. Ellipsis and stylistic variation in spoken and written English*. Harlow:
    Pearson Education.

Czech Student Subcorpus (CSS) (Non-native theses):
Huserek, J. 1999. *The indefinite article at the beginning of a sentence.* An unpublished thesis.
    Prague: Charles University, Faculty of Education, Department of English Language and
    Literature.
Kovačová, J. 1999. *Conversational analysis of children's and adult's corpora*. An unpublished
    thesis. Prague: Charles University, Faculty of Education, Department of English Language
    and Literature.
Šmilauerová, A. 2012. *TV sitcom* Friends: *analysis of character humor strategies based on the
    violation of Grice's conversational maxims.* An unpublished thesis. Prague: Charles
    University, Faculty of Education, Deparment of English Language and Literature.
Ženíšek, F. 2009. *Sociolinguistic concept of political correctness and its impact on modern English.*
    An unpublished thesis. Prague: Charles University, Faculty of Education, Department of
    English Language and Literature.

Czech Professional Subcorpus (CPS) (Non-native linguistic monographs):
Adam, M. 2006. *Functional macrofield perspective (A religious discourse analysis based on FSP).*
    Brno: Masarykova Univerzita, Pedagogická fakulta.
Povolná, R. 2003. *Spatial and temporal adverbials in English authentic face-to-face conversation.*
    Brno: Masarykova Univerzita, Pedagogická fakulta.
Tárnyiková, J. 2007. *Sentence complexes in text. Processing strategies in English and in Czech.*
    Olomouc: Univerzita Palackého v Olomouci, Filozofická fakulta
Urbanová, L. 2003. *On expressing meaning in English conversation. Semantic indeterminacy.* Brno:
    Masarykova Univerzita v Brně, Filozofická fakulta.

Secondary sources

Bachtin, M. M. 1980. *Román jako dialog.* Prague: Odeon.
Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan 1999. *Longman grammar of spoken and
    written English.* Harlow: Pearson Education.
Bhatia, V. K. 1993. Analysing genre: language use in professional settings. London: Longman.

Bjorkman, B. 2011. Pragmatic strategies in English as an academic lingua franca: ways of achieving communicative effectiveness. *Journal of Pragmatics*, *43* (4), 950–964.

Brendel, E., J. Meibauer & M. Steinbach (eds) 2011. *Understanding quotation*. Mouton Series in Pragmatics. Berlin: De Gruyter Mouton.

Čmejrková, S. 1996. Academic writing in Czech and English. In E. Ventola & A. Mauranen (eds) *Academic writing: intercultural and textual issues.* Amsterdam: John Benjamins, 137–152.

Dušková, L. et al. 1994. *Mluvnice současné angličtiny na pozadí češtiny*. Prague: Academia.

Ferenčík, M. 2012. Politeness aspects of EFL interaction: a discussion of a conversational encounter from the VOICE Corpus. *Prague Journal of English Studies*, *1* (1), 109–132.

Fludernik, M. 1993. *The fictions of language and the languages of fiction. The linguistic representation of speech and consciousness*. London: Routledge.

Greenbaum, S. & R. Quirk 1990. *A student's grammar of the English language*. Harlow: Longman.

Hoffmannová, J. 2008. The reproduction of one's own speech or the speech of others: from L. Doležel to contemporary communication and corpus research. In J. Králová & Z. Jettmarová (eds) *Tradition versus modernity. From the classic period of the Prague School to translation studies at the beginning of the 21st century*. Prague: FFUK, 101–123.

Huddleston, R. & G. Pullum 2002. The Cambridge grammar of the English language. Cambridge: Cambridge University Press.

Hyland, K. 2000. *Disciplinary discourses. Social interactions in academic writing*. Harlow: Pearson Education.

Hyland, K. 2006. *English for academic purposes. An advanced resource book*. London: Routledge.

Johansen, M. 2011. Agency and responsibility in reported speech. *Journal of Pragmatics*, *43* (11), 2845–2860.

Kaplan, R. B. 2000. Contrastive rhetoric and discourse analysis: who writes what to whom? When? In what circumstances? In S. Sarangi & M. Coulthard (eds) *Discourse and Social Life.* Harlow: Longman, 82–101.

Keizer, E. 2009. The interpersonal level in English: reported speech. *Linguistics*, *47* (4), 845–866.

Leech, G. N. & M. H. Short 1981. *Style in fiction*. Harlow: Longman.

Lipson, C. 2006. Cite right. *A quick guide to citation styles – MLA, APOA, Chicago, the sciences, professions, and more.* Chicago: The University of Chicago Press.

Mauranen, A., N. Hynninen & E. Ranta 2010. English as an academic lingua franca: the ELFA project. *English for Specific Purposes*, *29*, 183–190.

Pípalová, R. 2012. Framing direct speech: reporting clauses in a contrastive study. *Prague Journal of English Studies*, *1* (1), 75–107.

Povolná, R. & O. Dontcheva-Navrátilová (eds) 2012. *Discourse interpretation: approaches and applications.* Newcastle upon Tyne: Cambridge Scholars Publishing.

Quirk, R., S. Greenbaum, G. Leech & J. Svartvik 1985. *A comprehensive grammar of the English language.* Harlow: Longman.

Swales, J. 1990. *Genre analysis. English in academic and research settings.* Cambridge: Cambridge University Press.

APPENDIX 1.

Abbreviations and symbols

Abs. – absolute value
ARSA – anticipatory report of speech act
COMB – combined forms
CSS – Czech student subcorpus
IF – implicit frame
RP – Renata Pípalová
TOT – total

AIS – anticipatory indirect speech
CA – content anticipator
CPS – Czech professional subcorpus
EPS – English professional subcorpus
RC – reporting clause
SA – stance adverbial