A Corpus-Based Empirical Account of Adverbial Clauses Across Speech and Writing in Contemporary British English

Alex Chengyu Fang

Department of Chinese, Translation and Linguistics City University of Hong Kong Tat Chee Avenue, Kowloon, Hong Kong acfang@cityu.edu.hk

Abstract. Adverbial subordinators are an important index of different types of discourse and have been used, for example, in automatic text classification. This article reports an investigation of the use of adverbial clauses based on a corpus of contemporary British English. It demonstrates on the basis of empirical evidence that it is simply a misconceived notion that adverbial clauses are typically associated with informal, unplanned types of discourse and hence spoken English. The investigation initially examined samples from both spoken and written English, followed by a contrastive analysis of spontaneous and prepared speech, to be finally confirmed by evidence from a further experiment based on timed and untimed university essays. The three sets of experiments consistently produced empirical evidence which irrefutably suggests that, contrary to claims by previous studies, the proportion of adverbial clauses are a significant characteristic of planned, elaborated discourse.

1 Introduction

It is commonly accepted that adverbial clauses are registerially important, especially between speech and writing as two major modes of discourse. A recent consensus is that there are more adverbial clauses in speech than in writing. In his research on linguistic variations across speech and writing, Biber reports that "*that*-clauses, *WH*-clauses and adverbial subordinators co-occur frequently with interpersonal and reduced-content features such as first and second person pronouns, questions, contractions, hedges, and emphatics. These types of subordination occur frequently in spoken genres, both interactional (conversation) and informational (speeches), but they occur relatively infrequently in informational written genres" ([1], p230). More recently, this observation has been extended and introduced in the automatic analysis of biochemical text. In [2], Biber and Jones introduce a research approach that "combines corpus-linguistic and discourse-analytic perspectives to analyse the discourse patterns

T. Salakoski et al. (Eds.): FinTAL 2006, LNAI 4139, pp. 32-43, 2006.

[©] Springer-Verlag Berlin Heidelberg 2006

in a large corpus of biology research articles. The primary goals of the study are to identify vocabulary-based Discourse Units (DUs) using computational techniques, to describe the basic types of DUs in biology research articles as distinguished by their primary linguistic characteristics (using Multi-Dimentional analysis), to interpret those Discourse Unit Types in functional terms, and to then illustrate how the internal organization of a text can be described as a sequence of DUs, shifting among various Discourse Unit Types (p151).

Biber's claim is by no means unique. Thompson in [3] presents a similar claim, showing that the presence of subordination has to do with the formal/informal division and that, in terms of clause preference, speech appears to make use of more adverbial clauses and writing more non-finite clauses. More famously, in [4] and [5], Halliday observes that speech and writing are both complex systems but in different ways: speech is more complex in terms of sentence structures while writing in terms of high lexical density. In his opinion, the structural complex found in speech is characterised by a relatively higher degree of hypotaxis which involves subordination of various kinds such as adverbial clauses.

However, results of these past empirically based studies are far from conclusive. For one reason, they seem to have based their claims on either small samples or data that is not adequately defined or validated. In [1], for instance, it is not clear at all how many tokens of the spoken genre were used in the study. Instead, the basic figures were all normalised to a text length of 1,000 words. But even so, one easily questions the reliability of the data and indeed the validity of the analysis. For the mean frequencies of face-to-face conversations used in [1], as another example, the average number of infinitives per thousand tokens is as many as 13.8, far too high when compared with results of more recent studies such as [6], where infinitives account for fewer than 9 occurrences per thousand tokens in direct conversations. Indeed, [1] is based on frequencies collected from automatically analysed texts for its spoken and written samples. It is also worth pointing out that [1] makes use of the London-Lund corpus of English, which was produced over half a century ago.

This article reports an experiment that was aimed at a full review of the distribution of adverbial clauses across speech and writing. The experiment was performed on the basis of the understanding that conclusive results can only be obtained from first of all samples of authentic contemporary data and secondly from carefully designed analysis of the material that is manually validated and hence reliable. The next section will describe the data used in the experiment in terms of corpus composition and annotation.

2 Methodology

The methodology adopted in the current study was to investigate the distribution of different types of adverbial clauses across speech and writing based on a representative corpus of contemporary English. The scope of investigation would cover not only finite adverbial clauses but the non-finite ones, including infinitival, present participial and past participial constructions. The aim was to conclusively establish the differences in

the use of adverbial clauses, in frequential terms, across speech and writing. A second step would be to ascertain the variation of these clauses within the spoken and the written genres respectively.

Spoken				Written				
	Private				Student Writing			
	S1A1	direct conversations	90	be	W1A1	untimed essays	10	
	S1A2	distanced conversations	10	rinte	W1A2	timed essays	10	
	Public			L L	Correspondence			
ogue	S1B1	class lessons	20	ž	W1B1	social letters	15	
Dialo	S1B2	broadcast discussions	20		W1B2	business letters	15	
	S1B3	broadcast interviews	10		Informatio	onal		
	S1B4	parliamentary debates	10		W2A1	Learned: humanities	10	
	S1B5	legal cross-examinations	10		W2A2	Learned: social sciences	10	
	S1B6 business transactions]	W2A3	Learned: natural sciences	10	
	Unscripted				W2A4	Learned: technology	10	
	S2A1	spontaneous commentaries	20		W2B1	Popular: humanities	10	
	S2A2	unscripted speeches	30		W2B2	Popular: social sciences	10	
	S2A3	Demonstrations	10		W2B3	Popular: natural sciences	10	
logue	S2A4	legal presentations	10	rinted	W2B4	Popular: technology	10	
Mono	Mixed			ш.	W2C1	Press news reports	20	
	S2B1	broadcast news	20		Instructional			
	Scripted	l			W2D1	1 Administrative writing		
	S2B2	broadcast talks	20		W2D2	Skills and hobbies	10	
	S2B3	non-broadcast talks			Persuasive	e		
					W2E1 Press editorials		10	
					Creative			
				W2F1	Fiction	20		

Table 1. The	composition	of ICE-GB
--------------	-------------	-----------

The International Corpus of English (ICE) corpus was used in the current study as source of empirical evidence. The ICE project was launched by Professor Sidney Greenbaum at the Survey of English Usage, University College London. This project, participated by twenty national and regional teams, aims at the grammatical description of English in countries and regions where it is used either as a first or an official language ([6], p3). The British component of the corpus (ICE-GB) consists of 300 texts of transcribed speech and 200 texts of written samples, of 2,000 word tokens each, generally dated from the period 1990-1994. The component texts were selected according to registerial specifications. The spoken section, which contains 60% of the total corpus in terms of words, is divided between dialogues and monologues. The

dialogues range from private direct and distanced conversations to public situations such as broadcast discussions and parliamentary debates. The written samples are divided into two initial categories: non-printed and printed. The former is a collection of university essays and letters of correspondence. The latter has four major divisions: informational, instructional, persuasive, and creative. Table 1 presents an overview of the design of the corpus, with indications of text IDs, categories, and number of samples assigned to the category.

As can be seen from the corpus composition, ICE-GB provides an ideal setting for an empirical investigation of the variation in the use of adverbial clauses across speech and writing. First of all, the corpus is divided into spoken and written sections and thus allows for some general indications of distribution. Secondly, each major mode within the corpus contains genres that display a continuum between the spontaneous and the prepared, the informal and the formal, the timed and untimed, etc, thus allowing for the validation of hypothesis whether the use of adverbial clauses can be discussed along these lines, alongside the spoken-written division.



Fig. 1. The ICE parse tree for (1)

ICE-GB has been grammatically tagged, syntactically parsed and manually checked. The parsing scheme indicates a full analysis of the phrase structures and assigns syntactic functions to these constituents. Consider (1). (1) Electrical pulses travel from cell to cell, carrying messages which regulate all the body functions. <W2B-023-004>

This example in ICE-GB, taken from the fourth sentence in Text 23 of Genre W2B, receives the syntactic tree structure in Figure 1.

Each node in an ICE-GB tree comprises two labels: function and category. For example, SU NP() is interpreted as 'syntactic subject realised by the category NP or noun phrase'. Similarly, NPPR AJP(attru) indicates an attributive adjective phrase performing the function of an NP premodifer. The leaf nodes, i.e., the lexical items, are enclosed within curly brackets. As can be seen from Figure 1, Example (1) is analysed as a main clause consisting of a subject and a verb, with three adverbials: two realised by the prepositional phrases *from cell to cell* and one realised by a non-finite present participial clause *carrying messages which regulate all the body functions*. Features associated with the adverbial clause indicate that it does not have an overt subordinator (*zsub*), that its main verb is present participial (*ingp*), and that this clause does not have an overt subject (*-su*). The detailed annotation thus indicates explicitly the category names such as the clause and the phrase type as well as their syntactic functions such as subject and adverbial. ICE-GB therefore allows for unambiguous retrieval of different types of adverbial clauses.

3 The Experiments

The experiments examined the frequency distribution of finite adverbial clauses as well as the non-finite ones (infinitival, present participial, and past participial) in ICE-GB. There are three procedures. First, the experiment aimed to establish the overall distribution of adverbial clauses across the spoken and the written sections. Secondly, samples of spontaneous and prepared speech were examined to ascertain whether preparedness could be seen as a continuum of changes for the use of adverbial clauses. Finally, samples of timed and untimed university essays were used to validate the hypothesis that adverbial clauses also demonstrate a predictable variation as a function of degrees of preparedness in written English.

3.1 Uses of Adverbial Clauses Across Speech and Writing

As a first step, the complete corpus was used to obtain empirical indications of the different uses of adverbial clauses across speech and writing. Frequencies of occurrence were respectively collected from the spoken and the written sections of ICE-GB. The statistics include the total number of sentences and clauses in these two sections. Statistics were also collected for the total number of sentences involving the use of adverbial clauses and the exact number of adverbial clauses in these two sections. Two proportions were calculated: the total number of sentences with at least one adverbial clause over the total number of sentences, and the total number of adverbial clauses in the set total number of sentences. The former indicates the proportion of sentences in ICE-GB that make use of adverbial clauses. The latter shows

the proportion of adverbial clauses in the corpus since there often are multiple adverbial clauses in one sentence or utterance and it is useful to have such an indication. These two proportions thus indicate how often adverbial clauses are used and how complex the sentence structure is (assuming that structural complexity can be measured in terms of clause subordination). Table 2 summarises the results.

	S	poken	W	/ritten	Total		
	(59,470)		(2	4,084)	(83,554)		
	#	%	#	%	#	%	
Sentence	7124	11.98	6474	26.88	13598	13.27	
Clause	7809	13.13	7052	29.28	14861	17.79	

Table 2. Adverbial clauses across speech and writing

Initial results were simply contrary to what previous studies have suggested: the uses of adverbial clauses are more frequent in writing than in speech. As Table 2 clearly indicates, a much higher proportion of sentences in writing make use of adverbial clauses. To be exact, adverbial clauses are more than twice likely to occur in writing than in speech. In writing, 25.42% of the sentences make use of adverbial clauses in contrast to only 12.49% of the sentences with an adverbial clause in speech. The same difference can be observed in terms of the number of adverbial clauses: there are over 30 adverbial clauses per one hundred sentences in writing compared with fewer than 15 adverbial clauses per one hundred sentences in speech.¹

3.2 Types of Adverbial Clauses Across Speech and Writing

The distribution of different types of adverbial clauses was investigated in order to verify that the observed difference was not the result of a skewed use of any one particular type. The second experiment examined the distribution of finite adverbial clauses with an overt subordinator and the non-finite ones, which include infinitival, present participial and past participial adverbial clauses. They are illustrated respectively by examples (2)-(5) with the relevant sections underlined.

- (2) And I think the question is bigger than that <u>because it's from both sides</u>. <#S1A-001-054>
- (3) Having said that, I can really only say how it was for me when I came to work.
 <#S1A-001-056>
- (4) And you condemn the series <u>having seen a bit of one of them</u>. <#S1A-006-105>

¹ It makes more sense in terms of sentences rather than words. As a general guide, there are 600,000 words in the spoken section of the corpus and 400,000 words in the written section. In terms of words, therefore, there are 1.46 adverbial clauses per hundred words in speech, compared with 1.86 in writing.

(5) *The actual work surface was a very thick piece of wood, dumped on top, <u>all held</u> <u>in place by words</u>. <#S1A-009-200>*

The results are summarized in Table 3. As can be clearly seen, this second experiment also indicate that written samples of the ICE corpus make much more extensive use of the adverbial clause, be it finite, infinitival, or participial. The finite ones occur twice as many times in writing than in speech. For the other three types of adverbial clauses, the proportion for the written genre is even higher than for the spoken genre. Consider the infinitival clauses, for example. In writing, they are nearly three times more likely to be used than in spoken discourse (5.43% vs 1.98%), largely echoing previous observations that writing is characterised by a higher content of infinitives compared with spoken English (see, for example, [6] and [8]). This proportion is even greater with the other two types of non-finite adverbial clauses.

		Spoken (59,470)		W (2	/ritten 4.084)	Total (83,554)	
		#	%	#	%	#	%
A _{sub}	Sentence	5172	8.69	3954	16.42	9126	10.92
	Clause	5787	9.73	4430	18.39	10217	12.23
A _{infin}	Sentence	1122	1.89	1254	5.21	2376	2.84
	Clause	1177	1.98	1308	5.43	2485	2.97
Aing	Sentence	691	1.16	1023	4.25	1714	2.05
	Clause	704	1.18	1066	4.43	1770	2.12
A _{edp}	Sentence	139	0.23	243	1.01	382	0.46
	Clause	141	0.24	248	1.03	389	0.47
Total	Sentence	7124	11.98	6474	26.88	13598	16.27
	Clause	7809	13.13	7052	29.28	14861	17.79

Table 3. Types of adverbial clauses across speech and writing

We may incidentally note that past participial clauses are the least frequent type of adverbial clauses, with only 141 found in speech and 248 in writing in the whole corpus.

3.3 Types of Adverbial Clauses Across Spontaneous and Prepared Speech

Empirical indications thus irrefutably suggest that, contrary to previous claims, adverbial clauses are a marked characteristic of the written genre, in line with non-finite clauses that also characterise writing. However, to conclude that this difference in terms of use is due to different levels of elaboration, we need further empirical evidence. We need to prove that such variations can be observed not only across speech and writing, but also within the spoken and the written sections as a function of varying degrees of elaboration.

To this end, a sub-corpus of 180,000 words was created with S1A texts in ICE-GB,

representing spontaneous private conversations. A second sub-corpus was also created, this time with the first 40 texts in S2B, representing talks prepared and scripted for public broadcast. These two genres thus may be seen as forming a continuum between what was unprepared and what was carefully prepared, therefore a measure of different degrees of elaboration.

The results are summarised in Table 4, where we can read that, as an example, the subcorpus of spontaneous conversations contains a total number of 1,574 sentences that make use of finite adverbial clauses, accounting for 5.34% of the total number of sentences in the sub-corpus. On the other end of the continuum, as another example, we duly observe a higher proportion of finite adverbial clauses, that is, 12.81% in terms of sentences and 13.53% in terms of clauses. It is important to note that this general trend can be observed for all of the different types of adverbial clauses.

		Spontaneous (29,490)		Scripted (5 793)		Total (35,283)	
		#	%	#	%	#	%
•	Sentence	1574	5.34	742	12.81	2316	6.56
Asub	Clause	1757	5.96	784	13.53	2541	7.20
Δ	Sentence	271	0.92	253	4.37	524	1.49
Ainfin	Clause	279	0.95	260	4.49	539	1.53
۸.	Sentence	190	0.64	161	2.78	351	0.99
Aing	Clause	193	0.65	163	2.81	356	1.01
Δ.	Sentence	21	0.07	35	0.60	56	0.16
Aedp	Clause	21	0.07	36	0.62	57	0.16
Total	Sentence	2056	6.97	1191	20.56	3247	9.20
	Clause	2250	7.63	1243	21.46	3493	9.89

Table 4. Types of adverbial clauses across samples of spontaneous and scripted speech

It is thus reasonable to suggest that within speech the proportion of adverbial clauses increases as a function of degrees of elaboration, formality, and preparedness.

3.4 Types of Adverbial Clauses Across Timed and Untimed Essays

Having established that in speech the proportion of adverbial clauses is largely a function of elaboration or formality or preparedness, we want to do the same for the written samples. We want to argue, on empirical basis, that adverbial clauses not only mark a spoken-written division, that they also mark a continuum between what is spontaneous and what is scripted in speech, and that they also mark a degree of preparedness in writing.

Conveniently, the ICE-GB corpus contains a category coded W1A, which includes 20 texts evenly divided into two sets. Both sets were unpublished essays written by university students. The only difference is that the first set was written within a pre-designated period of time while the second set comprises samples written without

the time constraint. If the higher use of adverbial clauses were indeed the result of a higher degree of elaboration or preparedness, then we would observe more uses in the untimed set than in the timed set. This consideration led to a third experiment, whose results are summarised in Table 4.

		Timed		Untimed		Total		
		(1,057)		(1,046)	(2,103)		
		#	%	#	%	#	%	
Asub	Sentence	156	14.76	203	19.41	359	17.07	
	Clause	171	16.18	235	22.47	406	19.31	
۸	Sentence	62	5.87	61	5.83	123	5.85	
Ainfin	Clause	65	6.15	64	6.12	129	6.13	
Aing	Sentence	59	5.58	51	4.88	110	5.23	
	Clause	59	5.58	55	5.26	114	5.42	
A _{edp}	Sentence	10	0.94	16	1.53	26	1.23	
	Clause	10	0.94	16	1.53	26	1.23	
Total	Sentence	287	27.15	331	31.64	618	29.29	
	Clause	305	28.86	370	35.37	675	32.09	

Table 5. Types of adverbial clauses across samples of timed and untimed essays

Again, we duly observed a consistent increase in the proportion of adverbial clauses from one end of the continuum, timed essays, to the other end of the continuum, untimed essays. For instance, we observe that there are 16.18 finite adverbial clauses per 100 sentences for the timed essays. The untimed essays make more uses of finite adverbial clauses, 22.47 per 100 sentences. The same trend can be observed for all of the different types of adverbial clauses, except the infinitival ones. 62 sentences were observed to contain a total of 65 adverbial clauses in timed essays. In the untimed essays, 61 sentences were found to use a total of 64 infinitival adverbial clauses. While the differences are only marginal and can be dismissed as occasional, this group of texts will be examined in a future study for a possible relation between text types and uses of infinitival clauses.

For the purpose of the current study, it can be observed that in the untimed essays as a whole 31.64% of the sentences made use of adverbial clauses, almost 4.5% higher than 27.15% for the timed group. The results thus support the suggestion that within writing the proportion of adverbial clauses indicates different degrees of preparedness in terms of time.

3.5 Discussions

We have thus observed that, in the first place, adverbial clauses mark a division between spoken and written English in the sense that the spoken samples have a lower proportion of adverbial clauses than the written samples. This is true not only for finite adverbial clauses but non-finite ones, including infinitival, present participial and past participial constructions. Secondly, the experiments also produced empirical evidence that the frequency distribution of adverbial clauses follows a predictable and regular growth curve from spontaneous conversations to scripted public speeches. The same trend can be observed from within the written sample themselves, where the proportion of adverbial clauses in general increase from timed essays to untimed essays. As Figure 2 clearly demonstrates², the proportion of adverbial clauses per 100 sentences in ICE-GB consistently increases along a continuum between spontaneous conversations and untimed university essays. What is remarkably surprising is the fact that the occurrence of adverbial clauses in spontaneous conversations accounts for only about 7.5% of the utterances. What is equally surprising is that the occurrence of adverbial clauses in spontaneous for over 35% of the sentences, over 4.6 times as much as that in speech. The sharp contrast between speech and writing shown in Figure 2 argues strongly against the claims of past studies.

The graph also shows the average proportions of adverbial clauses in the two modes are nicely situated between the two sections within the same continuum. First of all, the average proportion of adverbial clauses in speech is shown in the figure to be between spontaneous conversations and scripted public speeches, suggesting a consistent increase in speech along the 'preparedness' register. In the written section of the continuum, the average proportion of adverbial clauses in writing rests between timed and untimed essays, again suggesting a consistent increase, continuing the trend from the spoken section, along the 'preparedness' register.

This is clear and irrefutable evidence that, contrary to results of previous studies, there are more adverbial clauses in writing than in speech, at least as far as contemporary British English is concerned and there is no obvious reason why other varieties of English should be seen otherwise. In the light of the evidence that the experiments came up with, observations such as the following is plainly not in line with what can be empirically observed in contemporary data: "Adverbial clauses appear to be an important device for indicating information relations in a text. Overall, Thompson (1984 [3]) and Biber (1988 [9]) find more adverbial clauses in speech than in writing." ([1], p235).

While it is evident from Figure 2 that speech and writing demonstrate a vast difference in terms of the use of adverbial clauses, it is clear at the same time that adverbial clauses are not as much a factor of speech vs writing division as a degree of preparedness in discourse. To be exact, it is acceptable to suggest on the basis of empirical evidence that degrees of information elaboration dictate the proportion of

- Spon: spontaneous conversations
- *Speech*: complete spoken samples
- Scripted: scripted broadcast news and talks
- *Timed*: timed university essays
- Writing: complete written samples
- *Untimed*: untimed university essays

 $^{^{2}}$ The X axis in Figure 2 has legends indicating the proportion of adverbial clauses in the following groups of samples in ICE-GB:



Fig. 2. The increase of adverbial clauses as a function of degrees of preparedness

adverbial clauses: the more elaborate the sample (defined in terms of preparedness), the more adverbial clauses, thus again contrary to a previous claim that '[t]he subordination features grouped on Factor 6 apparently mark informational elaboration that is produced under strict real-time constraints, resulting in a fragmented presentation of information accomplished by tacking on additional dependent clauses, rather than an integrated presentation that packs information into fewer constructions containing more high-content words and phrases" ([1]).

4 Conclusion

To conclude, this article reported an experiment to investigate the distribution of adverbial clauses across speech and writing. The experiment used ICE-GB, a corpus of contemporary British English that contains both transcribed speech and written samples. The detailed syntactic annotation of the corpus and manual validation of the analysis ensured that adverbial clauses could be accurately retrieved. These two features of the experiment are a clear advancement on past studies that made use of either old-fashioned data produced over half a century ago or unreliable analysis automatically performed by the computer without manual checking. The results irrefutably demonstrate that, contrary to claims by past studies, the proportion of adverbial clauses do not simply mark a division between the spoken and written genres. Empirical evidence strongly suggests that the proportion of adverbial clauses is also a function of varying degrees of preparedness, which can be independently demonstrated from within the spoken and written genres. It is thus reasonable to postulate that the spoken-written division is perhaps better perceived as a continuum of preparedness,

from spontaneous private conversations at one extreme to untimed carefully prepared writing at the other, along which the proportion of adverbial clauses consistently change in a predictable fashion.

Acknowlegement

The author would like to thank Professor M.A.K. Halliday for comments on an earlier draft of the article.

References

- 1. Biber, D. 1988. Variation across Speech and Writing. Cambridge: Cambridge University Press.
- Biber, D. and J. Jones. 2005. Merging corpus linguistic and discourse analytic research goals: Discourse units in biology research articles. In *Corpus Linguistics and Linguistic Theory 1-2* (2005). pp 151-182.
- Thompson, S. 1984. Subordination in Formal and Informal Discourse. In D. Schffrin (ed), *Meaning, Form, and Use in Context: Linguistic Applications*. Washington DC: Georgetown University Press. pp 85-94.
- 4. Halliday, M.A.K. 1979. Differences between Spoken and Written Language: Some Implications for Literacy Teaching. In G. Page, J. Elkins, B. O'Connor (eds), *Communication* through Reading: Proceedings of the Fourth Australian Reading Conference, Brisbane, 25-27 August 1978, Vol. 2, Diverse Needs: Creative Approaches. Australian Reading Association. Pp 37-52.
- 5. Halliday, M.A.K. 1985. Spoken and Written Language. Victoria: Keakin University Press.
- 6. Fang, A.C. 1995. The Distribution of Infinitives of Contemporary British English: A Study Based on the British ICE Corpus. In *Oxford Literary and Linguistic Computing*, 10:4. pp 247-257.
- 7. Greenbaum, S. (ed) 1996. Comparing English World Wide: The International Corpus of English. Oxford: Oxford University Press.
- 8. Mair, C. 1990. *Infinitival Complement Clauses in English*. Cambridge: Cambridge University Press.
- 9. Biber, D. 1988. Adverbial stance types in English. In Discourse Processes 11. pp 1-34.