Lingue e Linguaggi
Lingue Linguaggi 9 (2013), 21-40
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v9p21
http://siba-ese.unisalento.it, © 2013 Università del Salento

COHESION AS INTERACTION IN ELF SPOKEN DISCOURSE

THOMAS CHRISTIANSEN
UNIVERSITÀ DEL SALENTO

Abstract – Hitherto, most research into cohesion has concentrated on texts (usually written) only in standard Native Speaker English – e.g. Halliday and Hasan (1976). By contrast, following on the work in anaphora of such scholars as Reinhart (1983) and Cornish (1999), Christiansen (2011) describes cohesion as an interactive process focusing on the link between text cohesion and discourse coherence. Such a consideration of cohesion from the perspective of discourse (i.e. the process of which text is the product -- Widdowson 1984, p. 100) is especially relevant within a lingua franca context as the issue of different variations of ELF and inter-cultural concerns (Guido 2008) add extra dimensions to the complex multi-code interaction. In this case study, six extracts of transcripts (approximately 1000 words each), taken from the VOICE corpus (2011) of conference question and answer sessions (spoken interaction) set in multicultural university contexts are analysed in depth by means of a qualitative method.

Keywords: Anaphora, Cohesion, Discourse, ELF, Interaction.

1. Introduction

In this paper,[1] we analyse an area which has hitherto largely been overlooked in the field of English as a Lingua Franca (ELF), namely cohesion, although much valuable work has been done on the pronunciation, lexico-grammatical and pragmatic aspects of ELF (see, among many others, Jenkins 2000, 2002, Mauranen and Ranta, 2009, Cogo and Dewey 2012). A corpus made up of six extracts taken from the VOICE corpus (2011) is used to examine the types of cohesion produced in the corpus as a whole and to compare the output of different selected speakers with the aim of examining how speaker's linguistic background, namely their respective L1s, affect the type of cohesion that they produce in ELF.

Cohesion, the way that items in a text presuppose each other – that is, rely upon each other for their interpretation – is an important area of language, first discussed in depth by Halliday and Hasan (1976) although Quirk et al. (1972), had previously dedicated a chapter to the same subject. Halliday and Hasan (1976) and subsequent studies (e.g. Halliday 2004) have concentrated on presupposition only within texts (usually written) and then only in standard English. However a wider view that looks at supposition in the context of discourse (of which the text is just one component) has been taken in a field related to cohesion, so-called anaphora (see Reinhart, 1983 and Cornish, 1999), which owes less to Halliday and Hasan (1976), and has concentrated on the highly complex relations between pro-forms (e.g. pronouns) and full forms (the noun phrases which "resolve" them or serve as their antecedents).

[1] This paper was originally read at the ELF5 conference (English as a Lingua Franca), Boğaziçi University, (Istanbul, Turkey), 24-26 May 2012. A much shortened version of this paper comprising parts of Sections 2, 3 and Subsection 4,2 appears as Christiansen (forthcoming).

Christiansen (2011) describes cohesion as an interactive process focusing on the link between text cohesion and discourse coherence, this latter concept seen itself as not something immutable but dependant on whose perspective is taken: the addressor or addressee. Viewing cohesion within the broader, and more complicated, context of discourse (seen as the process of which text is the product see Widdowson 1984, p. 100) is especially relevant within a lingua franca context because the issue of the existence of what are in effect different varieties of ELF, as well as wider inter-cultural concerns, not least the different perceptions and expectations of participants from different lingua-cultural backgrounds (see Guido 2008). Such factors add extra dimensions to the complex multi-code interaction of which cohesion must be seen not only as the key element in the co-construction of a dialogic text in interaction, but crucially as constituting the interface between the various ELF varieties in the ongoing development of discourse.

Cohesion then plays a central role in how participants interpret discourse and also in how the same discourse is constructed. Studies of ELF discourse cannot afford to ignore this aspect of language. Unfortunately, as things stand, work on cohesion has been limited almost exclusively to NS (native speaker) discourse and even there, studies are still in their early stages (see Christiansen 2011). In ELF, there is obviously a need for parallel studies to those on NS discourse to be attempted and for examination of just how far existing concepts and terminology gleaned from analysis of NS discourse are applicable or useful.

This present studies aims to start that process by attempting to catalogue, categorise and count the various types of cohesive device essentially as identified by Halliday and Hasan (1976) but with some minor revisions, more in the interest of facilitating analysis (see Section 3.), in such a way that future studies and theories regarding cohesion in ELF discourse may be constructed around such raw data taken directly from ELF discourse seen not as a poor cousin to 'proper' NS discourse or as an imperfect imitation of it but as a bona fide phenomenon in its own right that should be studied on its own terms.

As regards organisation of this paper, in Section 2, we briefly discuss the corpus then in Section 3 outline our method of analysis and very briefly discuss those parts of our terminology that differ to that introduced in the seminal work of Halliday and Hasan (1976), which still constitutes the 'conventional wisdom' in the field. In Section 4 we discuss our results, dedicating the first subsection 4.1. to an analysis of general types of cohesion found in the corpus as a whole and subsection 4.2. to a comparison of the types of cohesion produced by six individual speakers of different L1 backgrounds.

2. Corpus

The corpus used for this analysis was made up of six extracts (approximately 1,000 words each – taken from the beginning of each),[2] taken from the VOICE corpus (2011) of conference question and answer sessions. Basic data relating to the corpus is given in Table 1:

[2] The reason that the extracts were not exactly 1,000 words each is that we choose as a break-off point the first change in speaker turn after the 1,000 word mark.

PRqas 18 (1048 words)	S1	S2	S3	SX
Speaker's L1* ▶	Dut	Nor	Fin	?
% Discourse ▶	16.52	81.95	0.38	1.15

PRqas 19 (1094 words)	S1	S2	S3	SS	SX
Speaker's L1* ▶	Span	Kor	Eng	?	?
% Discourse ▶	8.14	89.39	2.2	0.09	0.18

PRqas 224 (1280 words)	S1	S4	S5	S6	SX
Speaker's L1* ▶	Ger	Rus	Hun	Spa	?
% Discourse ▶	5.86	34.14	42.50	17.42	0.08

PRqas 407 (1031 words)	S1	S2	S3	S4	S5
Speaker's L1* ▶	Ger	?	Slv	Ger	Cze
% Discourse ▶	8.24	14.45	33.85	26.19	17.26

PRqas 409 (1182 words)	S1	S2	S3	S4	S5	SX6	SXm
Speaker's L1* ▶	Ger	Slv	Slo	Cze	?	?	?
% Discourse ▶	5,84	16.24	63.28	11.34	3.13	0.08	0.08

PRqas 495 (1058 words)	S1	S2	S3	S4	S13	SS	SX-4	SX-m
Speaker's L1* ▶	Spa	Eng	Chi	Spa	Spa	?	?	?
% Discourse ▶	8.60	5.58	82.04	2.74	0.19	0.09	0.38	0.38

* Key: Chi = Chinese; Cze = Czech; Dut = Dutch; Eng = English; Fin = Finnish; Ger = German; Hun = Hungarian; Kor = Korean; Nor = Norwegian; Rus = Russian; Slo = Slovakian; Slv = Slovene; Spa = Spanish.

Table 1
Composition of corpus.

In the top left hand corner the VOICE corpus identification code for each transcript is given; in brackets the approximate number of words in the extract is given (approximate because the figure includes some vocal sounds such as 'um', 'er', or laughter). To the right along the top row are the VOICE identification codes for each speaker in the extracts. On the bottom row is the percentage of the discourse which that particular speaker produces. On graphs on the left, give a representation of how the discourse of each extract is divided among the contributing speakers and allows, at a glance, to see that PRqas19 is the extract

where one speaker is most dominant and PRqas407 that where discourse is most evenly shared between the various participants.

3. Method of analysis

Because of the many and varied ties that may exist between items within a discourse, it is difficult to arrive at a simple and comprehensive classification of the cohesion in any particular texts, and an analysis may run to several lists of different kinds of cohesive items, each several pages long. Lexical cohesion in particular is problematic as alternative classifications exists (cf. Halliday and Hasan 1976, Hasan 1979, 1984, Halliday 2004), not all of them based entirely on objective criteria (see Christiansen 2011).

In this analysis, we use elements of Halliday and Hasan's (1976) description for grammatical ties (substitution, ellipsis and conjunction), and for lexical ties, (Hoey 1991) because, though not perfect, it is, of the available classifications, the most rigorous and easiest to be apply consistently (see Christiansen 2011) (Table 2):

Simple Repetition	Complex Repetition
Items of same word class sharing same lexical morpheme (*bear*/*bears*)	1) Same word class but with no common morpheme (*am*; *is*). 2) Like simple, share a common morpheme but are of different word class (*drug* noun / *drugging* verb*)* 3) Antonyms containing the same lexical morpheme (*happy/unhappy)*
Simple Paraphrase	Complex Paraphrase
When one item can replace another text without change of meaning, or having to undergo any transformation (e.g. *produce*/*cause*). If the substitution is reciprocal, the paraphrase is *Mutual*, if not, *Partial.*	Like *Simple*, but one item includes another, but shares no common lexical morpheme (e.g. *hot*/*cold*; *author*/*writings*).

Table 2
Hoey's (1991) categorisation of types of lexical repetition (Christiansen 2011, p. 274).

As regards semantic ties and in particular what Halliday and Hasan problematically term *reference*, we use Christiansen's 2009 terminology for types of noun phrases based on the means of reference (in the conventional philosophic sense) which they constitute: names (proper nouns), epithets (noun phrases headed by common nouns) and so-called deictic devices for pronouns and determiners in general (Table 3):

Means of Reference		Specific Kind of NERE	Syntactic Manifestation
Representation (*Rep NERE**)	Describing	*Epithet*	noun phrase headed by common noun
	Labelling	*Name*	proper noun
Deixis/Indication (*Deictic NERE**)		*Deictic Device*	pronouns/ possessive determiners

* Key: NERE = Nominal Entity Referring Expression

Table 3
Different means of reference and corresponding types of NERE (Christiansen 2009b, p. 36)

A quantitative method was employed whereby the corpus was analysed manually from different perspectives according to each kind of cohesive device, the results for the individual analyses given and compared in the following section.

4. Results

In the subsections below we will examine the results for the various analyses that we carried out on the corpus. In subsection 4.1 we will look at general results for the extracts taken as a whole, in 4.2 on data for a selection individual speakers of different L1 backgrounds.

4.1. General results, whole corpus

In this section we will look at general results for the corpus as a whole, breaking down this analysis into separate sections for anaphora, co-reference, substitution and ellipsis (4.1.1), conjunction (4.1.2) and lexical cohesion (4.1.3).

4.1.1 Anaphora, co-reference, substitution and ellipsis

Looking first at general results for anaphoric reference, co-reference[3] (the two being closely linked in that both phenomena typically co-occur in identity chains),[4] substitution and ellipsis. Examples of each general category taken from the corpus are highlighted below (1-4):

> (1) but **the norwegian er students** [full form] **they** [anaphor] just say hi and pass? (PRqas18)
>
> (2) how does it look with the TOtal debts of the of **the countries** [co-referring expression 1] especially (.) i heard that there are big debts (.) of of **the countries** [co-referring expression 2] in i- in in the PENsion reforms. (PRqas407)
>
> (3) it's not always (hard for us) because you (.) VERY often (need) two **countries** you know? **country** of background and the **country** where you are WORKING or maybe a third **one** [nominal substitute]? (PRqas18)
>
> (4) this won't be enough for all the **questions** but (.) which is erm t- erm yes er i wanted to (.) to ask you **two** [nominal ellipsis of head noun *questions*] in particular (PRqas224)

Anaphoric reference (1) can come in three main forms: personal (the pronouns and determiners *I*, *you*, *me*, *her* etc.); demonstrative (the pronouns, determiners and adverbs referring to space and distance *this*, *those* etc.);[5] and comparative (pronouns, determiners and adverbs that draw comparison with some other item *other*, *same*, *more* etc.). Co-reference (2) can be realised by a variety of lexical forms including, as here, same item reiteration

[3] Anaphora and co-reference are closely linked the difference being that anaphora refers a variety of asymmetrical relations between items, where one item (e.g. pronoun, substitute) is dependent on a full-form for interpretation.

[4] Strings of expressions that share the same referent (see Hasan1984; Christiansen 2009).

[5] The definite article *the* also falls into this group. In this study it was ignored as a separate phenomena as it was dealt with under the category of def epithet (see Table 7).

and the use of synonyms and paraphrases provided that the referent of the various forms is the same, which is not the case with the repetitions of *country* in Example 3. Substitution (3) and ellipsis (4), two closely related phenomena, can both be divided into nominal, verbal and clausal.

General figures for analysis of these phenomena are given in Table 4 below.

	TOTAL	MEAN
Anaphor comp ref*	21	3.5
Anaphor dem ref*	53	8.83
Anaphor per ref*	168	28
Cataphor per ref*	6	1
Co-ref def epithet*	103	17.17
Co-ref epithet*	82	13.67
Co-ref name*	37	6.17
Co-ref proposition*	3	0.5
Substitution clausal	1	0.17
Substitution nom 'one'*	2	0.33
Ellipsis clausal	11	1.83
Ellipsis nom*	4	0.67
Total anaphor	248	41.34
Total co-ref*	225	37.5
Total substitution	3	0.5
Total ellipsis	15	2.5

	Mean distance between ties (words)	Mean no. of ties in identity chain	No. of different identity chains
PRqas18	39.02	4.03	35
PRqas19	46.54	5.07	15
PRqas224	37.34	3.19	42
PRqas407	83.54	3.19	32
PRqas409	58.54	3.48	33
PRqas495	51.85	3.86	28
MEAN	52.81	3.80	30.83
ST.DEV*	15.53	0.65	8.23

*Key: comp = comparative; co-ref = co-reference; def = definite; dem = demonstrative; nom = nominal; per = personal; ref = reference; ST DEV = standard deviation.

Table 4
General figures anaphor, co-reference, substitution and ellipsis in whole corpus

The bottom left section of Table 4 shows that anaphor and co-ref most common form of cohesion followed (by far) by ellipsis and then substitution. The top right section shows that in PRqas407 (the extract where discourse is most evenly distributed – see Table 1) the mean distance in words between ties in the same identity chain is largest by far. The greatest number of ties is found in PRqas409. In PRqas224, where two speakers dominate (34.14% and 42.5% each – see Table 1) there is the highest number of chains. By contrast, in PRqas19, where one speaker predominates, 89.39% – see Table 1) the number of different chains is notably lower than every other extract. Of the three parameters: distance / number of ties / number of chains, it is the second where standard deviation between the figures for the difference is lowest, indicating greatest uniformity here.

To examine the interactive aspect of ties, we categorised ties according to whether the previous item in the identity chain was produced by the same speaker within the same turn (SSST); by the same speaker in a different (i.e. previous) turn (SSDT); or by a different speaker (DS). The results for this analysis are presented in Table 5, the figures given being occurrences expressed as percentages of the total for that type of item.

	DS	SSDT	SSST
Anaphor comp ref %	14.29	4.76	80.95
Anaphor dem ref %	13.21	16.98	69.81
Anaphor per ref %	2.38	8.93	88.69
Cataphor dem ref %	0	0	100
Cataphor per ref %	0	0	100
Co-ref def epithet %	18.63	16.67	64.71
Co-ref epithet %	15.85	12.20	71.95
Co-ref name %	32.43	13.51	54.05
Co-ref proposition %	0	0	100
Substitute clausal %	0	0	100
Substitute nom 'one' %	0	0	100
Ellipsis clausal %	23.08	15.38	61.54
Ellipsis nom %	7.69	0	15.38

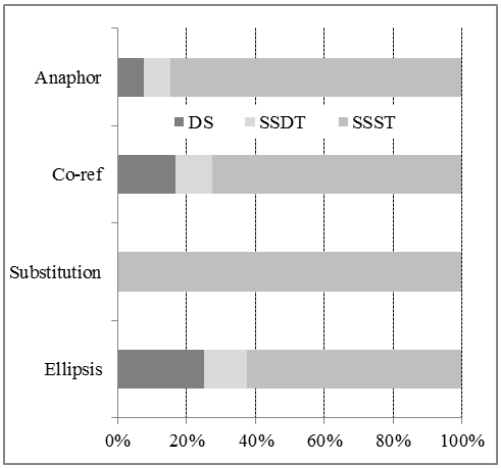

Table 5
Anaphoric reference, co-reference, substitution and ellipsis within and between speaker turns

As is clearly shown on the graph in the right-hand column all of the relatively few cases of substitution found in the corpus (see Table 4) were produced by the same speaker in the same speaker turn. Generally it can be seen most cases of anaphora are also SSST but relatively fewer cases of co-reference are, which ties in with the well-documented fact that considerations of saliency (including proximity) influence type of noun phrase selected in an identity chain (see Cornish 1999, Christiansen 2009). It is with ellipsis that SSST is least common (but still accounts for the majority of cases), with nominal ellipsis indeed there are only cases of SSST and DS (respectively 33.33% and 66.67%). This however is not too surprising given the fact that clausal ellipsis in particular is often used in replies (e.g. "maybe you should speak slowly?" – "yeah [I should speak more slowly]").

In the next table (6), we look at the same data as in Table 5 but from the perspective of proportion of ties within and between turns in each extract, the figures given being expressed as percentages of the total for that kind of tie for that extract.

	PRqas18 %	PRqas19 %	PRqas224 %	PRqas407 %	PRqas409 %	PRqas495 %	MEAN %	ST. DEV
DS*	3.74	11.48	16.67	24.64	15.85	8.64	13.5	6.62
SSDT*	14.95	3.28	10	5.8	0	28.4	10.4	9.35
SSST*	81.31	85.25	73.33	69.57	84.15	62.96	76.09	8.15

* Key: DS = Different Speaker; SSDT = Same Speaker Different Turn; SSST = Same Speaker Same Turn

Table 6
Proportion of ties within and between ties in each extract

It is notable that the mean for SSST is much higher than for DS or SSDT and the standard deviations for all three categories are similar: that for DS lowest (indicating most uniformity between extracts), and that for SSDT marginally highest.

Looking at the individual extracts, it is interesting to note that the figure for DS is highest in extract PRqas407, which is also the extract where the discourse is distributed most evenly among the participants (see Table 1). By contrast, SSST is highest for

PRqas19 (the extract where discourse is most dominated by a single participant). The highest percentage of SSDT is highest in PRqas18 (also a relatively one-sided discourse). These results are indicative of the fact that where a discourse is made up of contributions by different speakers, there is more scope for ties between turns (both by the same and different speakers). Below (4.2), we examine this point further from the perspective of individual speakers.

A final consideration when looking at components of identity is the kind of antecedent triggers by which the concept referred to (the antecedent) is accessed in the mind (see Cornish 1999). A simple chain may be made up of a full noun phrase (the trigger) followed by a string of pronouns (anaphors).[6] In Table 7, figures for the occurrences of various kinds of trigger are given:

	PRqas18	PRqas19	PRqas224	PRqas407	PRqas409	PRqas495	MEAN
Comp ref	1	1	0	0	0	0	0.33
Per ref	3	0	2	0	0	1	1
Def epithet	4	7	12	6	11	5	7.5
Epithet	16	3	13	9	10	11	10.33
Name	2	2	2	8	4	2	3.33
Proposition	7	2	0	10	8	0	4.5
Verb	1	0	0	0	0	0	0.17
Deictic	4	1	2	0	0	1	1.33
NERE	22	12	27	23	25	18	21.17
Verb / Proposition	8	2	0	10	8	0	4.67

Table 7
Anaphoric reference / Co-reference / Substitution – Ellipsis: Types of Antecedent Trigger

As Table 7 shows, there is notable difference between extracts in the kinds of triggers which occur. For example, propositions (clauses) are the most common antecedent triggers for anaphors of various kinds in PRqas407, yet do not feature at all in PRqas224 and PRqas495.Overall, as could be expected, NEREs (nominal entity referring expressions, i.e. full noun phrases: see Table 3) are most common type of trigger followed by verb / propositions (e.g. "I will *help* you – and then students think *it*'s about everything")[7] and finally by deictics (pronouns – see Table 3), which are normally used anaphorically rather than as triggers in the third person (see Christiansen 2009), and first and second deictics were excluded from our analysis as they are typically used deictically or ostentively[8] to refer to something (e.g. the addressor , the addressee) in the physical context rather than in the text or discourse. The fact that third person deictics function as triggers at all, as in such cases the anaphora is not endophoric (intra-textual) but exophoric (extra-textual but still intra-discoursal).[9] In this way, the antecedent is left implicit (as in the example of a

[6] As Christiansen (2009a, 2009b, 2011), documents, this is often not the case; full noun phrases may be used elsewhere in the chains, among other considerations, or reasons of referential efficacy or for informativeness.
[7] PRqas18.
[8] See Lyons (1999: 160) on the distinction.
[9] Following Cornish's (1999) definition of these terms rather than Halliday and Hasan's (1976) – see Christiansen (2011).

verb trigger quoted showing a degree of sophistication at the cohesive level). Indeed, on occasion there is a mismatch between the deictic device and the full-form trigger "*this er investments* are getting *its* importance".[10] Such a phenomenon should not be dismissed as mere anomalies as examples of apparently mismatched or unmatched deictics are common in natural spontaneous discourse (e.g. "Oh I was on the bus + *he* didn't stop at the right stop"),[11] where they may be exophoric or function as triggers for some implicit but mutually manifest[12] antecedent (as in the last example). Indeed the fact that they can be used without any breakdown in communication is testimony to the fact that behind the text there exists a shared discourse accessible to all the participants even in the context of ELF where most, if not all, of the participants are non-native speakers (see Table 1), and use different varieties of ELF.[13]

4.1.2. Conjunctions

Conjunctions involve a specific type of cohesion in the own right whereby clauses or clause complexes are linked to other clauses in explicit ways indicating the way in which the two are related as propositions. The kinds of relationships that conjunctions can express may be categorised in various ways. Halliday and Hasan (1976) list five (together with intonation, which we shall not discuss here): additive (*and*, *furthermore* etc); adversative (*but*, *notwithstanding this* etc.); causal (*so*, *because* of this etc.); temporal (*then*, *after that* etc.); or continuative (*well*, *of course* etc.). By contrast, Halliday (2004) adopts three general categories which are applicable also to other intra- and interclausal bonding devices such as modification in word phrases, namely: elaboration (specification), extension (addition) and enhancement (qualification).

As our examples show, conjunctions may be manifested not just by the word class of conjunctions and coordinators in traditional grammars but also by longer expressions (see Halliday and Hasan 1976, Halliday 2004, Christiansen 2011).

In Table 8, we give the general figures for the occurrence of conjunctions in the corpus. Again a great deal of diversity can be seen in type and frequency of conjunctions in various extracts. Continuatives, which constitute what others have called discourse markers,[14] for example are used markedly more in PRqas 224 and PRqas495 than others. For Halliday and Hasan (1976), additives and continuatives are the most frequent. With Halliday's revision (2004) the differences between extracts remain, but there is a marked preference for extension (but this is the largest category in terms of variety and number of items that it includes).

[10] PRqas407

[11] Brown and Yule (1983, p. 219)

[12] This concept is central to Sperber and Wilson's concept of relevance (1987) which they argue lies at the heart of all communication.

[13] It has been suggested to us that we might have also made a distinction in terms of level of so-called linguistic competence (see for example Council of Europe 2001). We avoided such an approach as it classes participants in the extracts merely as 'learners' and demotes the status of the discourse itself to a mere imitation of some NS equivalent (see our comments in Section 1)

[14] See for example: Louwerse and Mitchell (2003), Bazzanella *et al.* (2007).

	Halliday and Hasan (1976)					Halliday (2004)		
	ADD	ADVERS*	CAUSAL	TEMP	CONTIN*	ELAB*	ENHAN*	EXTEN*
PRqas18	28	8	5	6	1	1	11	35
PRqas19	10	4	8	3	4	1	11	13
PRqas224	18	6	9	4	47	5	11	22
PRqas407	27	9	12	5	10	8	14	28
PRqas409	14	7	9	5	10	6	14	15
PRqas495	16	13	20	9	42	14	29	15
MEAN	18.83	7.83	10.5	5.33	19	5.83	15	21.33
% Total	30.62	12.74	17.07	8.67	30.89	14	35	54

* Key: ADVERS = Adversative; CONTIN = Continuative; ELAB = Elaboration; ENHAN = Enhancement; EXTEN = Extension

Table 8
Frequency of different general types of conjunctions in extracts according to categorisations of Halliday and Hasan (1976) and Halliday (2004).

In Table 9, figures are given for items within categories, in the second column on the left an example is given of each type.

It transpires from Table 9 that the greatest variety of subtypes of conjunctions are found with the general categories of additive and adversative (five subtypes each) and temporal (four). Causal has only one subtype, which is indeed expressed by a single form *so* and not by any alternatives (e.g. *hence*, *thus*, *therefore*) are found at all in the corpus, notwithstanding the fact that this single example constitutes 17.07% of the conjunctions found in the entire corpus (see Table 8). In a departure from Halliday and Hasan's original analysis, we have added the subcategory of "quasi-continuatives". This includes items which are not among the six specific examples listed by Halliday and Hasan as what we have called "pure" continuatives (i.e. *now*; *of course*; *well*; *anyway*; *surely* and *after all*) but which do match the description of a continuative as an conjunctive device that, in an albeit vague way, indicates that something has gone before (see Christiansen 2011) and show the characteristics of discourse markers in general. Examples of these are *you know*, *I don't know*, *yes/yeah/yah*, *like*, *sort of*, *I suppose*, *okay*.

In Table 10, we look at the same data as in Table 9 from the perspective of proportion of ties between turns (i.e. ignoring SSST – same speaker same turn) in each extract, the figures given being expressed as percentages of the total for that kind of conjunction for that extract.

		PRqas18	PRqas19	PRqas224	PRqas407	PRqas409	PRqas495	MEAN
Ext/int* simple additive	*and*	27	8	11	17	8	9	13.33
Ext/int* simple alternative	*or*	0	1	5	0	1	0	1.17
Int* apposition exemplificatory	*for example*	1	1	0	5	3	3	2.17
Int* apposition expository	*I mean*	0	0	2	2	1	4	1.5
Int* complex emphatic additive	*furthermore*	0	0	0	3	1	0	0.67
ADDITIVE (TOTAL)		28	10	18	27	14	16	18.83
Ext/int* containing and	*but*	8	3	5	8	4	5	5.5
Ext* contrastive simple	*and*	0	0	0	1	1	1	0.5
Ext/int* dismissal open-ended	*anyway*	0	0	1	0	0	0	0.17
Ext/int* emphatic	*however*	0	1	0	0	0	0	0.17
Int* contrastive	*in fact*	0	0	0	0	2	7	1.5
ADVERSATIVE (TOTAL)		8	4	6	9	7	13	7.83
Ext/int* general simple	*so*	5	8	9	12	9	20	10.5
CAUSAL (TOTAL)		5	8	9	12	9	20	10.5
Ext* simple preceeding	*before that*	2	0	0	0	0	0	0.33
Ext* complex repetitive	*again*	0	1	2	2	1	1	1.17
Ext* simple sequential	*then*	4	2	1	3	4	7	3.5
Int* correlative sequential	*first*	0	0	1	0	0	1	0.33
TEMPORAL (TOTAL)		6	3	4	5	5	9	5.33
"Pure" continuative	*well*	1	2	6	3	6	1	3.17
"Quasi" continuative	*you know*	0	2	41	7	4	41	15.83
CONTINUATIVE (TOTAL)		1	4	47	10	10	42	19

* Key: Ext = External; Int = Internal

Table 9
Frequency of different subtypes of conjunctions in extracts according to categorisations of Halliday and Hasan (1976)

		PSqas18 %	PRqas19 %	PRqas224 %	PRqas407 %	PRqas409 %	PRqas495 %	MEAN %
DS*	Adversative ext/int containing 'and'	0	0	0	3.03	0	0	0.51
	Causal ext/int general simple	0	25	0	0	0	25	8.33
	Temporal int correlative sequential	0	0	0	0	0	50	8.33
	Continuative	0	0.88	5.26	3.51	2.63	4.39	2.78
SSDT*	Additive ext/int simple additive	3.75	1.25	0	0	0	0	0.83
	Additive int apposition exemplificatory	0	0	0	0	0	7.69	1.28
	Causal ext/int general simple	1.59	0	0	0	0	4.76	1.06
	Continuative	0	0	0	0	0	1.75	0.29

* Key: DS = Different Speaker; SSDT = Same Speaker Different Turn; SSST = Same Speaker Same Turn

Table 10
Proportion of ties between ties in each extract

Looking at conjunctive ties between turns, either by the same or by a different speaker (SSDT and SD respectively), it can be seen that for DS continuatives are the most widespread (occurring in all extracts except PRqas18) while the mean is highest for causal external / internal general simple (i.e. *so*) and temporal internal correlative sequential (e.g. *first*) which however have a high concentration only in certain extracts, the former in PRqas19 and PRqas495, the latter only in PRqas495.

Turning to SSDT, none of the types of conjunctions are particularly widespread, additive external / internal simple additive (e.g. *and*) and causal external / internal general (i.e. *so*) appear in only two extracts each, and in neither is the concentration high (between 1.25 and 4.76). the means are also low (between 0.29 and 1.28).

These results are indicative of the fact that those conjunctions that link propositions in different turns tend to do so either with something immediately preceding (i.e. another speaker's turn) or, if to a turn by the same speaker, to one which has been interrupted briefly by another speaker, thus constituting a resumption. The reason for the high proportion of continuatives with DS is because expressions such as *well*, *now*, *anyway*, *you know*, *I don't know*, *yeah* are often used as fillers at the beginnings of speaker turns especially in replies to direct or indirect questions.

4.1.3. Lexical Cohesion

Above in Section 3, we note how lexical cohesion has proved to be an area where revision of Halliday and Hasan's original categorisation has been order. On Table 2 above, we present the classification presented by Hoey (1991), which has proved to be more reliable for objective analysis. In Table 11, we give the general results (given as percentages) for lexical cohesive ties found in the corpus:

	PRqas18%	PRqas19%	PRqas224%	PRqas407%	PRqas495%	PRqas409%	MEAN	ST.DEV
SLR*	57.8	51.50	68.59	40.51	47.75	37.05	50.53	11.57
CLR*	24.77	22.89	21.47	9.17	18.02	6.93	17.21	7.47
SMP*	1.83	0	0	0	0.9	0.45	0.53	0.73
SPP*	4.89	19.07	2.24	32.62	14.19	32.08	17.52	13.01
CP*	10.70	6.54	7.69	17.7	19.14	23.49	14.21	6.87

* Key: SLR = Simple Lexical Repetition; CLR = Complex Lexical Repetition; SMP = Simple Mutual Paraphrase; SPP = Simple Partial Paraphrase; CP = Complex Paraphrase

Table 11
Lexical cohesive ties in extracts

As can be seen from Table 11, there is great variety between extracts. For example, for PRqas19, SLR makes up 57.8% of the lexical cohesive ties, while for PRqas409, it is only 37.05%. The standard deviation (ST.DEV: right-most column) shows that there is greatest uniformity between the various extracts in the frequency of simple mutual paraphrase and least with simple partial paraphrase. As can be seen from the mean figures, the most common type of lexical cohesive tie is SLR, then SPP.

Looking at the figures for lexical chains (groups of items which are bound together by lexical ties of the various kinds listed by Hoey 1991), [15] in Table 12 we give the general figures for all types of lexical cohesive device in the various extracts:

	PRqas18	PRqas19	PRqas224	PRqas407	PRqas495	PRqas409	MEAN	ST.DEV%*
Total no. of lexical items in chains	327	367	312	469	444	664	430.5	5.05
No. different lexical chains	82	96	94	77	99	119	94.5	4.36
Mean no. lexical items per chain	3.99	3.82	3.32	6.09	4.48	6.21	4.65	2.6

* Key: ST.DEV% = standard deviation expressed as a percentage of the total for all categories (thereby allowing for direct comparison between different rows).

Table 12
General figures for all types of lexical cohesive device in extracts

As regards number of items in lexical chains, PRqas409 has the highest numbers of items that are found in chains. By contrast PRqas224 has fewest. The relatively high figure for standard deviation (right-most column) confirms that there is a fairly wide difference in this respect between categories.

Looking at the number of different chains in each extract, there is a greater degree of uniformity between the extracts than for number of lexical items found in chains (see standard deviation). PRqas409 is the extract with the largest number of different lexical chains and PRqas407 that with the lowest.

[15] These may include some items that also feature in identity chains (e.g. epithets see 1.4.1), but at this (lexical as opposed to semantic) level of analysis, whether or not they are co-referential is beside the point.

Again it is PRqas409, where the mean number of items per chain is highest, and PRqas224 where it is lowest. The discernibly low standard deviation (only 2.6) shows that there is a great deal of uniformity between the various extracts.

This data shows that while there are differences in the types of lexical cohesion found in each extract (see Table 11) there is notable similarity when it comes to number of lexical chains and the number of different lexical items, of whatever kind, within them. Lexical cohesion is achieved differently in the various extracts but the amount of lexical cohesion and number of lexical chains in each is largely uniform.

4.2. Results for a selection of Individual Speakers

In this section, we shall look at the results for individual speakers of different L1s in an attempt to examine how far the type of cohesion produced by each follows or diverges from the general patterns identified in section 4.1.

For each extract we isolated the discourse produced by the dominant speaker as identified on Table 2, i.e. he or she who produced the largest proportion of the discourse. In Table 13, we give the general figures (given as percentages) regarding different types of cohesion produced by each speaker (identified in the top row, with their respective L1s given in brackets):

	PRqas 18 % S2 (Nor)*	PRqas19 % S2 (Kor)*	PRqas224% S5 (Hun)*	PRqas407% S3 (Slv)*	PRqas409% S3 (Slo)*	PRqas495% S3 (Chi)*	MEAN	ST.DEV
Anaphor	12.8	6.95	9.33	5.31	4.95	6.84	7.69	2.94
Substitution	0.17	0.25	0	0	0	0	0.07	0.11
Ellipsis	0	0	1.04	0.41	0	0.37	0.3	0.41
Conjunction	7	6.2	18.13	12.65	5.15	17.38	11.09	5.79
Lexical	80.03	86.6	71.5	81.63	89.9	75.42	80.85	6.83

* Key: Chi = Chinese; Hun = Hungarian; Kor = Korean; Nor = Norwegian; Slo = Slovakian; Slv = Slovene.

Table 13
Comparison of percentage of different types of cohesion
produced by one selected speaker for each extract.

Table 13 shows the figures for speakers of various L1s ranging across Europe and Asia, from Norway to Korea, including Slavic languages such as Slovene and Slovakian, Uralic (Hungarian), Germanic (Norwegian), Sino-Tibetan (Chinese) and disputed Altaic / isolate (Korean). The figures are given as percentages of the total for that speaker (e.g. 12.8% of the cohesive ties produced by the sample speaker from PRqas18 S2 are anaphors). Without exception, lexical cohesive ties account for the largest part by far of the cohesion produced, even though the standard deviation is the highest of all the categories indicating least uniformity here in the actual percentage.

The lowest percentages overall and across the board (see the very low standard deviation indicating uniformity) are taken up by substitution and ellipsis. This is interesting because although the low incidence of substitution may be put down to the fact that the rules governing it may be less accessible to non-native speakers as it, would seem most peculiar to English[16] (see Christiansen 2011), ellipsis, resting as it does on a basic princi-

[16] It might once more be tempting the view the issue in terms of linguistic competence (see fn above), which at one level is a factor. This would however in our view be missing the larger more interesting picture of

ple of abbreviation and reduction of given items, would seem to be resource available to all languages and thus constitute a universal as one can suppose do anaphor, conjunctions and lexical cohesive ties. However, ellipsis and substitution both operate at the grammatical level and thus are intricately bound to the complexities, or vagaries as they may sometimes seem to a non-native speaker, of the morphosyntax of English, it is therefore fair to assume that non-native speakers could not easily transfer competences from their L1s to assist them in these specific areas of cohesion.

It is also interesting that while the mean for anaphora, a recognised language universal (see Haegeman 1991, Cornish 1999), is relatively high (7.69), the standard deviation is also relatively low (2.94) indicating again a high degree of uniformity.

Conjunctions are more frequent than anaphors but there is markedly less uniformity (see standard deviation). Conjunctions also constitute universals but it is worth noting that intonation – of kinds not tagged in VOICE – can play a role here (see Halliday and Hasan 1976, p. 271) so our data here may be incomplete and it is conceivable that the true figure is higher. It is certainly interesting to note that non-native ELF speakers of different L1s do avail themselves of this resource more readily than anaphora. This may indicate that conjunctions, of certain kinds at least, are more fundamental cohesive devices than anaphors and more easy to handle again given the fact that deictic devices in particular have to be imbedded within the syntax of sentences whereas conjunctions can be slotted relatively easily between propositions with little concern for syntax or concord / agreement (and see our observations in 4.1.1 about apparent mismatch between anaphors and triggers). Furthermore, as Christiansen (forthcoming) notes, conjunctions do not merely express inherent logical relations between propositions but can also help shape those relations allowing the speaker to manipulate them for reasons of stance (Jaffe 2009).[17] In this way, as Christiansen (2011) also observes conjunctive force can be equated with illocutionary force and conjunctions can fulfill an interpersonal function.

Some possible evidence for the fact that speaker's L1 does have an effect on the nature of the cohesion that they produce is given by the fact that, of the selected speakers, the figures for the L1 Slovene and L1 Slovakian are most similar (columns 5 and 6), both Slavic languages, although there is not enough data here to establish whether this is not in fact merely a coincidence.

In Table 14, we look more closely at general results for anaphoric reference, co-reference, substitution and ellipsis, figures given as percentages of the total. From Table 14, it transpires that only anaphora (comparative, demonstrative and personal reference) and co-reference (epithet: definite and non-definite and name) are used by all speakers and the other categories are only used by one speaker each (two in the case of clausal ellipsis). There overall picture then is of a degree of uniformity in the case of anaphora and co-reference but much less so with the other categories. This would underline that anaphora (excluding the subcategory of cataphora where the anaphor precedes the trigger in the text) and co-reference (except for when the trigger is a proposition – see 4.1.1) are the most basic kinds of cohesion in ELF contexts according to this selection of speakers.

Looking at conjunctions, in Table 15, we examine figures for conjunctions (continues overleaf).

how ELF speakers construct discourse along new lines and not by merely copying, consciously or not, native speaker models.

[17] Similarly, Gotti (2003, p. 107) notes that conjunctions have a pragmatic function: "which clarifies the purpose of the sentence that follows".

	PRqas 18 % S2 (Nor)	PRqas19 % S2 (Kor)	PRqas224% S5 (Hun)	PRqas407% S3 (Slv)	PRqas409% S3 (Slo)	PRqas495% S3 (Chi)
Anaphor comp ref*	2.04	3.33	11.76	9.68	6.45	2.67
Anaphor dem ref*	7.14	8.33	8.82	6.45	6.45	17.33
Anaphor per ref*	63.27	35	32.35	25.81	25.81	29.33
Cataphor per ref*	4.08	0	0	0	0	0
Cor ref def epithet*	13.27	18.33	11.76	35.48	22.58	20
Co-ref epithet*	6.12	16.67	23.53	3.23	27.42	21.33
Co-ref name*	3.06	16.67	5.88	16.13	9.68	6.67
Co-ref proposition*	0	0	0	0	1.61	0
Ellipsis verbal*	0	0	0	3.23	0	0
Ellipsis clausal*	0	0	5.88	0	0	2.67
Substitute nom*	0	1.67	0	0	0	0
Substitute clausal*	1.02	0	0	0	0	0

* Key: comp = comparative; co-ref = co-reference; def = definite; dem = demonstrative; nom = nominal; per = personal; ref = reference; ST DEV = standard deviation

Table 14
Anaphoric reference, co-reference, substitution and ellipsis in selected speakers' turns

	PRqas18 S2 (Nor)	PRqas19 S2 (Kor)	PRqas224 S5 (Hun)	PRqas407 S3 (Slv)	PRqas409 S3 (Slo)	PRqas495 S3 (Chi)
Additive ext/int simple additive	23	7	3	8	3	8
Additive ext/int simple alternative	0	1	0	0	1	0
Additive int apposition exemplificatory	1	1	0	3	2	3
Additive int apposition exemplificatory	1	0	0	2	0	4
Additive int complex emphatic additive	0	0	0	1	0	0
Additive ext/int dismissal open-ended	0	0	1	0	0	0
Adversative ext/int containing 'and'	6	3	3	3	4	5
Adversative ext contrastive simple	0	0	0	0	0	1
Adversative int contrastive	0	0	0	1	2	7
Adversative ext/int emphatic	0	1	0	0	0	0
Causal ext/int general simple	4	8	6	7	6	19
Temporal ext complex repetitive	0	1	0	0	0	1
Temporal ext correlative sequential	0	0	0	0	1	2
Temporal ext preceeding	1	0	0	0	0	0
Temporal ext simple sequential	4	0	0	0	3	4
Temporal ext simple sequential	0	2	0	0	0	0
Temporal int correlative sequential	0	0	0	0	0	1
Continuative "proper"	1	0	1	3	3	1
Quasi continuative	0	1	21	3	0	38
Number of different conjunctions	8	9	6	9	9	13

Table 15
Types of conjunction in selected speakers' turns.

Table 15 shows that there is also great diversity in frequency and type of conjunction used by different ELF speakers, with only additive external / internal simple additive (e.g. *and*), adversative external / internal containing 'and' (e.g. *but*), and causal external / internal general simple (i.e. *so*) being used by all of the selected speakers.

The variety of items used collectively by the different speakers shows how, even in ELF discourse, conjunctions can link propositions in myriad and subtly different ways.

The bottom rows shows that the variety of different conjunctions used by the various selected speakers ranges from six to thirteen (mean: nine). This shows again that flexibility with conjunctions is a common trait with all of the selected speakers, underlining the fact that conjunctions would seem to be a basic and versatile cohesive device regardless of speaker and their linguistic background in the ELF discourse examined here.

Turning to the last kind of cohesion, lexical, in Table 16 we show the various types of lexical cohesion found in the discourse of the selected speakers (figures given as percentages):

	PRqas18% S2 (Nor)	PRqas19% S2 (Kor)	PRqas224% S5 (Hun)	PRqas407% S3 (Slv)	PRqas409% S3 (Slo)	PRqas495% S3 (Chi)	MEAN	ST.DEV
SLR*	35.82	49	70.29	37	34.4	46.81	45.55	13.55
CLR*	16.	22.06	23.19	5.5	7.34	19.12	15.57	7.51
SMP*	0.85	0	0	0	0.46	0.98	0.38	0.45
SPP*	1.71	23.78	2.17	30	33.72	14.46	17.64	13.8
CP*	45.42	5.16	4.35	27.5	24.08	18.63	20.86	15.37

* Key: SLR = Simple Lexical Repetition; CLR = Complex Lexical Repetition; SMP = Simple Mutual Paraphrase; SPP = Simple Partial Paraphrase; CP = Complex Paraphrase

Table 16
Types of lexical cohesion in selected speakers' turns

Again with lexical cohesion (and as with extracts in general – see Table 11), there are differences between the types of lexical tie which are used by each selected speaker. It can be seen that overall and also for each speaker, except PRqas18 S2, SLR accounts for the largest percentage of lexical ties. The lowest category both overall and for each selected speaker is SMP. The scarcity of SMP is consistent with the high incidence of SLR as it shows that speakers tend to repeat the same forms rather than indulge in what Christiansen (2011) calls *avoidance of formal repetition*, which is one of the four principal factors that he identifies in noun phrase selection in written discourse, at least.

The figures for the other kinds of lexical cohesion are far less uniform (see standard deviation) except for SMP and CLR. As with conjunctions (see Table 15), the general picture is that different speakers all make use of this type of tie but in markedly different ways. Differently to conjunctions, which as we have argued above have an interpersonal function and in part serve to model the discourse in the way that reflects the speaker's stance, lexis is in large part determined by the ideational content of the discourse and thus is something over which the speaker has less control.

5. Conclusions

The ELF discourse studied here taken as a whole shows patterns of cohesion that can be described in terms of categorizations designed primarily for NS varieties of English. Modes of cohesion more particular to English, such as ellipsis and especially substitution, are less frequent indicating that ELF speakers make use of cohesive strategies transferred from their L1s. These universals however prove adequate in ELF to ensure that the discourse is cohesive.

It has been shown that individual ELF speakers from a variety of L1s, all dominant in their respective extracts of the discourse, achieve cohesion in markedly different ways especially within categories of cohesive device (see Tables 14-16). This leads to the hypothesis that speakers either avail themselves of cohesive devices that are transferred from their L1s or develop new ones specifically for use in ELF discourse.

This is a research question which merits a great deal of research in its own right and before speculating further than we have done here there is a need for more studies such as this to further ascertain not just how far conventional descriptions of cohesion from NS discourse fit ELF discourse, but also into aspects of cohesion in Languages other than English (see Christiansen 2009) to provide not only new perspectives but also better understanding of the cohesive strategies that ELF users bring with them from their own L1 and to investigate whether some cohesive strategies employed in ELF are not entirely new or hybrid and come neither from NS English nor from the speaker's L1. In the latter case, the presumption would be that there exists a set of basic linguistic universals that pertain to cohesion in discourse regardless of the form in which the text is encoded.

It would also be interesting to examine how NS speakers of English use cohesive devices in ELF discourse, in particular whether they adopt compensation strategies. Specific studies into substitution and ellipsis might be particularly appropriate to this aim.

References

Bazzanella, C, Garcea, A, Bosco, C, Gili Fivela, B, Miecznikowski J, and Tini Brunozzi, F. 2007, *Italian allora, French alors: functions, convergences, and divergences* in *Catalan Journal of Linguistics*, special issue, Cuenca M.J, *Contrastive Perspectives on Discourse Markers*, Universitat Autònoma de Barcelona, Servei de Publicacions, Bellaterra, *6*, pp. 9-30.

Christiansen, T. (forthcoming) *Cohesion as interaction in ELF spoken discourse: an analysis of question-answer sessions in university contexts*, in *Proceedings ELF 5*, Boğaziçi University, (Istanbul, Turkey), 24-26 May, 2012, (http://www.elf5.org/index.php?p=7).

Christiansen, T. 2009a, *Identity Chains and Noun Phrase Selection: A Case Study of Italian*, Berlin, VDM Verlag Dr Müller.

Christiansen, T. 2009b, *Designation and description: the informative function of referring expressions*, in *Lingue e Linguaggi* 2/2008, pp. 33-57.

Christiansen, T. 2011, *Cohesion: A Discourse Perspective*, Bern, Peter Lang.

Cogo, A. and Dewey M. 2012, *Analysing English as a Lingua Franca: A corpus-based investigation*, London, Continuum.

Cornish, F. 1999, *Anaphora, Discourse and Understanding*, Oxford, Oxford University Press.

Council of Europe 2001, *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge, Cambridge University Press.

Gotti, M. 2003, *Specialized Discourse*, Bern, Peter Lang.

Guido, M.G. 2008, *English as a lingua franca in cross-cultural immigration domains*, Bern, Peter Lang.

Haegeman, L. 1991, *Introduction to Government and Binding Theory*, Oxford, Blackwell.

Halliday, M.A.K. 2004, *An Introduction to Functional Grammar* (3rd edition revised by C.M.I.M. Matthiessen), London, Edward Arnold.

Halliday, M.A.K. and Hasan, R. 1976, *Cohesion in English*, London, Longman.

Hasan, R. 1979, *On the notion of text*, in Petöfi, J. S, *Text vs. Sentence: Basic Questions of Text Linguistics*, Hamburg, Helmut Buske.

Hasan, R. 1984, *Coherence and cohesive harmony*, in Flood, J, *Understanding Reading Comprehension*, Delaware, International Reading Association, pp. 181-219.

Hoey, M. 1991, *Patterns of Lexis in Text*, Oxford, Oxford University Press.

Jenkins, J. 2000, *The Phonology of English as an International Language*, Oxford, Oxford University Press.

Jenkins, J. 2002, *A sociolinguistically based, empirically researched pronunciation syllabus for English as an International Language*, in *Applied Linguistics* 23/I, pp. 83-103.

Louwerse, M.M. and Mitchell, H.H, 2003, *Towards a taxonomy of a set of discourse markers in dialog: a theoretical and computational linguistic account*, *Discourse Processes*, 35, pp. 199-239.

Mauranen, A. and Ranta, E, 2009, *English as a Lingua Franca: Studies and findings*, Newcastle upon Tyne, Cambridge Scholars Press.

Quirk, R, Greenbaum, S, Leech, G, and Svartvik, J, 1972, *A Grammar of Contemporary English*, London, Longman.

Reinhart, T. 1983, *Anaphora and Semantic Interpretation*, London, Croom Helm.

Sperber, D. and Wilson, D. 1986, *Relevance, Communication and Cognition*, Oxford, Blackwell.

VOICE 2011, *The Vienna-Oxford International Corpus of English* (version 1.1 online), Director: Barbara Seidlhofer; Researchers: Angelika Breiteneder, Theresa Klimpfinger, Stefan Majewski, Ruth Osimk, Marie-LuisePitzl. http://voice.univie.ac.at (13/11/11).

Widdowson, H.G. 1984, *Explorations in Applied Linguistics 2*, Oxford, Oxford University Press.