Qf Lingüístics

Lexical cohesion revisited.
A combined corpus and systemic-functional analysis

Revisió del concepte de cohesió lèxica. Una anàlisi de corpus i sistemicofuncional combinada

María de los Ángeles Gómez González

Universidad of Santiago de Compostela. mdelosangeles.gomez@usc.es

Abstract: In this chapter I argue for a refinement of the classic SFL approach to lexical cohesion. First, a literature overview is provided in which key principles and related categories are examined. In addition, the connection of cohesion and discourse coherence is addressed and an overview is provided of the wide range of applications that the former has in such fields as genre studies, language teaching and learning, psycholinguistics and computational linguistics, among others. The core SFL models of cohesion are then revisited in order to propose a modified taxonomy of lexical cohesion, involving five distinct types (*repetition*, *synonymy*, *opposition*, *inclusion* and *association*) that are cross-classified by four kinds of distance-based ties (*immediate, immediate-mediated, remote* and *remote-mediated*). After this, the model is attested in the telephone conversation and broadcast discussion components of the *International Corpus of English-Great Britain.* The analysis of 10,042 cohesive ties reveals that repetition and association are the most frequently used lexical cohesion strategies across the two genres. They are overwhelmingly produced over speakers' turns and remote-mediated ties. The results further indicate that lexical patterns collaborate in topic management, staging and turn-taking strategies.

Keywords: lexical cohesion; repetition; association; coherence; systemic-functional grammar.

Resum: En aquest capítol proposem una reorientació de l'aproximació clàssica de la lingüística sistèmica funcional (LSF) a la cohesió lèxica. En primer lloc, es presenta un panorama global de la bibliografia en el qual s'examinen els principis bàsics i les categories amb què es relaciona. A més, s'aborda la relació entre la cohesió i la coherència discursiva, i es fa una descripció de l'ampli espectre d'aplicacions que té la cohesió en camps com ara els estudis de gènere, l'ensenyament i l'aprenentatge de llengües,

la psicolingüística i la lingüística computacional, entre altres. Es revisen els models centrals de cohesió de la LSF amb la intenció de proposar una taxonomia modificada de la cohesió lèxica que inclou cinc classes diferents (*repetició, sinonímia, oposició, inclusió* i *associació*) que es tornen a classificar creuant-les amb quatre tipus de relacions basades en la distància (immediata, immediata-mediata, remota i remota-mediata). A continuació, el model se posa a prova tenint en compte les seccions de l'*International Corpus of English-Great Britain* corresponents als gèneres conversa telefònica i debat de radiodifusió. L'anàlisi de 10.042 relacions cohesives posa de manifest que la repetició i l'associació són les estratègies de cohesió més utilitzades en els dos gèneres. Es produeixen profusament en els torns de paraula dels parlants i en les relacions remotes-mediates. Així mateix, els resultats mostren que els patrons lèxics afavoreixen la gestió del tema discursiu, la posada en escena i les estratègies de presa de torn.

Paraules clau: cohesió lèxica; repetició; associació; coherència; gramàtica sistèmica funcional.

1. Introduction: Cohesion and related categories

Much of the relevant literature on cohesion is indebted to Halliday & Hasan's (1976) seminal book and its developments within Systemic Functional Linguistics (SFL), often regarded as "the best known and most detailed model of cohesion available" (Baker, 1992: 180; Moreno 2003; Xi 2010). The notion rests on the intuition that a text "hangs together" by a network of logical and semantic, anaphoric and cataphoric, relationships between a presupposing element (e.g. a pronoun) and the presupposed (e.g. its antecedent), which together establish a *cohesive tie* (underlined).

Cohesive ties can be classified into three groups, according to the criteria of *coreferentiality*, *co-classification*, and *coextension* (Halliday & Hasan 1985: 73). Coreferentiality is concerned with identity of reference between the members of a tie, as in the case of *a little nut tree* and *it* in (1), while *silver* and *golden* illustrate coextension because they refer to the field of meaning of 'metal'. In contrast, co-classification appears in (2) because *play the cello* and *does* refer to different members of identical classes, if it is assumed that wife and husband play and have different cellos.

(1)   I had <u>a little nut tree</u>. Nothing would <u>it</u> bear. But a *silver* nutmeg. And a *golden* pear.

(2)   I <u>play the cello</u>. My husband <u>does</u>, too. (Halliday & Hasan, 1985/1989: 73)

In addition, cohesive ties are categorized in terms of the distance between the elements. Thus we can speak of *immediate ties*, when the presupposed and the cohesive element are found in subsequent sentences, as in the tie *Alice-she* in [1] and [2], *remote ties*, when the distance between the two elements is extensive, as shown in the tie *Alice rubbed her eyes-Rub as she would* in [3] and [7], and *mediated ties*, when the presupposed element is found earlier in the discourse, but it has been followed through the relationship with some other element, as illustrated by *she* in sentence [5] which refers back to *Alice* in [3], but this tie is mediated by another instance of *she* in [4].

(3)   The last word ended in a long bleat, so like a sheep that <u>Alice</u> quite started [1]. <u>She</u> looked at the Queen, who seemed to have suddenly

wrapped herself up in wool [2]. Alice <u>rubbed</u> her eyes, and looked again [3]. <u>She</u> couldn't make out what had happened at all [4]. Was <u>she</u> in a shop [5]? And was that really–was it really a *sheep* that was sitting on the other side of the counter [6]? <u>Rub</u> as she would, she could make nothing more of it [7]. (Halliday & Hasan, 1976: 330)

The concept of mediated tie leads to that of *cohesive chain*, which occurs when a cohesive element refers back to an element that is itself cohesive with a still earlier element, and so forth (Halliday & Hasan, 1976). Cohesive chains can be of two different kinds: *identity* and *similarity*. Identity chains are made up of cohesive ties that share the same referent(s) (pronominals, reiterations, or equivalents), whereas similarity chains refer to elements that are nonidentical and are realised by co-classification or coextension. A crucial factor to the coherence of texts is *cohesive harmony* or cohesive chain interaction, which in turn is key to recognising *texture*, that is, the property that distinguishes texts from non-texts (Martin 2001). There is a minimum requirement that at least two members of one chain should stand in the same relation to two members of another chain (Halliday & Hasan, 1985: 91).

The notion of cohesion has been associated with that of coherence. The most widespread view is that coherence and cohesion reflect the opposition that exists between the external (or situational) and the internal aspects of texts, respectively. Hence, broadly while cohesion is viewed as an inter-sentential property of a text which is achieved through the grammatical and lexical text-internal devices that serve a cohesive function, coherence refers to a cognitive process of conceptual connectivity in which speakers construct a mental representation of the information conveyed in discourse comprehension against the background of shared and specific world knowledge. The underlying rationale is that, although they are often intertwined and usually work collaboratively, cohesion is neither a necessary nor a sufficient condition for discourse coherence. This is illustrated in (4), which shows that coherence can exist with few or no explicit markers of cohesion, and (5), a made-up example where the repetitions of *goldmine*(*s*) and *shape(s)*, all instances of lexical cohesion, do not achieve coherence:

(4)   A: <u>That</u>'ll be the phone
       B: I'm in the bath
       A: OK (Widdowson, 1979: 29)

(5)   The capital city of Santiago de Compostela is a <u>goldmine</u> of historic significance. <u>Goldmines</u> around the world come in all <u>shapes</u> and sizes. Note the <u>shape</u> of this lesion, located contouring the contact point. I don't like the <u>shape</u> of this hat.

A myriad of studies corroborate that cohesive devices contribute to building up discourse coherence and relevance such as the work by Moreno (2003), Sanders and Maat (2006), Taboada and Mann (2006), and Dontcheva-Navratilova and Povolná (2009), to mention but a few. In addition, other studies have explored the functionalities of specific cohesive makers (e.g. anaphoric or abstract nouns, repetition, discourse markers such as *well*, *but* and others) (Ochs Keenan, 1977; Tadros, 1985; Francis, 1986; Norrick, 1987; Schmid, 2000; Fraser, 2005; Tannen, 2007; Flowerdew, 2010). Likewise, many investigations have scrutinized how cohesive strategies vary and contribute to creating identities across different discourse types (e.g. monologic, dialogic, multimodal) and genres (e.g. narratives, academic language, legal texts, fund-raising letters, news reports) (Gutwinski, 1976; Tanskanen, 2006; Carter-Thomas, 2008; Flowerdew & Mahlberg, 2009; Gómez González, 2010, 2011, 2013), as well as across different languages and cultural backgrounds (Taboada, 2004; Yankova, 2006; Kunz, 2015; Steiner, 2015). Lastly, cohesion has also been used as a tool for topic detection and tracking (Stokes, 2004), as a measure of text readability in psycholinguistics and computational linguistics (CohMetrix), as a means of effective communication in (first, foreign, and/or second language) language education, or as a test to evaluate levels of faithfulness to the original text in translation studies (McNamara *et al.*, 2002; Ebrahimpourtaher & Eissaei, 2013; Struthers *et al.*, 2013).

In what follows further details are offered as to how the theory of cohesion has been developed within SFL in order to present and attest a combined corpus and SFL analysis of lexical cohesion.

2. A combined corpus and Systemic-Functional analysis of lexical cohesion

In SFL two kinds of cohesion are recognized, depending on the linguistic strategies used: *grammatical* or *lexical*, as shown in Table 1 (Halliday & Hasan, 1976). Grammatical cohesion traces cohesive ties that

are triggered by grammatical items (e.g. pronouns, determiners, conjunctive elements and the like) in *reference*, *substitution*, *ellipsis*, and *conjunction* as described in SFL.

Grammatical cohesion	*Lexical cohesion*
Reference	Reiteration
- personal	- same item
- demonstrative	- synonym
- comparative	- superordinate
Substitution	- subordinate
- nominal	- general word
- verbal	Collocation
- clausal	
Ellipsis	
- nominal	
- verbal	
- clausal	

Table 1. Cohesion in SFL: Halliday & Hasan's (1976) model

Reference (anaphora) is illustrated in (6), where the item *they* refer to *three blind mice* in the previous sentence. Likewise, *the same* in (7) is a substitute of *two poached eggs on toast*, whereas there is an instance of verbal ellipsis in the answer the verbal group in (8), and an illustration of conjunction in (9), where the conjunctive item *after* involves a presupposition of sequence between two clauses implying that what is expressed in the second one preceded what is expressed in the first.

(6) Three blind mice, three blind mice. See how they run! See how they run! (Halliday & Hasan, 1976: 31)

(7) I'll have two poached eggs on toast, please. I'll have the same. (Halliday & Hasan, 1976: 105).

(8) Have you been swimming? –Yes, I have. (Halliday & Hasan, 1976: 167).

(9)   He left the post office <u>after</u> he posted the letter. (Halliday & Hasan, 1976: 222).

Lexical cohesion, on the other hand, is achieved by the selection of vocabulary or such lexical items as nouns, verbs, adjectives and adverbs including multi-word units (e.g. *kick the bucket*) as shown in (10), where the antonymous adverbs *fast-slow* bear a relation of semantic contrast.

(10)   <u>Fast</u> learners learn quickly. <u>Slow</u> learners take their time. (Halliday & Hasan, 1976: 274)

Table 2 displays the core SFL models of lexical cohesion that have been couched in SFL, namely Halliday and Hasan (1976, 1985), Halliday (1985), (1994), Martin (2001), Halliday and Matthiessen (2014).

The models in Table 2 above underpin a common *system-oriented* perspective and a two-way distinction between *collocation* (i.e. words that are associated or co-occur in texts such as *candle-flame-flicker* or *hair-comb-curl-wave*) and *reiteration* (i.e. two occurrences of the same referent), including repetition (*girl-girl*), (near-)synonymy (*clothed-dressed*), antonymy (*fast-slow*), superordinates (*daisy, azalea, daffodil-flower* (superordinate)), meronyms (*tree-trunk, branch, leaf* (co-meronyms)) and general nouns (*fact, thing, situation*).

In this chapter I propose a modified approach to lexical cohesion (see also Gómez González, 2010, 2011, 2013). My point of departure is that, in order to obtain a functionally adequate usage-based account of lexical cohesion, a *discourse-specific* view, rather than a system-oriented perspective, should be adopted assuming that lexical meaning is (con)text-specific (Hoey, 1991; Morris & Hirst, 1991; McCarthy, 1988; Tanskanen, 2006). What this means is that the meaning potential of cohesion cannot be determined in isolation. Instead, the communicative potential of lexical ties must be determined against the context of occurrence, taking into account such factors as processing relevance and cognitive coherence in relation to the context-specific communicative needs of discourse participants, thereby emphasising the collaborative dimension of (lexical) cohesion (Stubbs, 2001).

Secondly, and as a corollary of the above, lexical cohesion has a discourse dimension that motivates its use and therefore must be taken

HALLIDAY & HASAN (1976)		HALLIDAY (1994) MARTIN (2001) HALLIDAY & MATTHIESSEN (2014)			HALLIDAY & HASAN: HASAN (1985)		
Reiteration	Repetition	Repetition			Repetition		
	(Near) Synonym	Synonymy	Same referent	Synonymy	Synonymy		
	Superordinate			Superordinate			
	General word		Different referent	Hyponymy	Antonymy		
				Meronymy			
		Antonymy			Hyponymy	Superordinate	
						Hyponym	
Collocation		Collocation			Meronymy		

Table 2. Models of lexical cohesion couched in SFL

on board in any effort to describe it. We shall see how lexical cohesion contributes to constructing both *global discourse topics* (e.g. cognitive schemes that compress the topic of a whole text into a single proposition) and *local topics* (e.g. Subject and Object participants that build up text-level topics) in different *topic continuity strategies* such as *reintroducing*, *shifting* (or *switching*) and *drifting* (Stenström, 1994: 151-162; Downing, 2000). Topic continuity can be identified whenever there are no signs of topic change and is generally characterized by the presence of referential continuity, temporal continuity, and/or specific markers that signal continuity (e.g. *and*, *moreover*, *besides*). Topic re-introduction, on the other hand, occurs when a given referent is taken up again after having been abandoned over a number of textual spans. Shifting involves a distinct movement to a related aspect of the current topic framework that is signalled by explicit devices such as framing expressions, intonation or backchannels. Lastly, drifting represents an almost imperceptible move to a related aspect of the current topic by means of either the partial variation of previously established topics, or otherwise by the association of concepts, ideas or events that bear some sort of cognitive relationship.

Thirdly, the proposed model in Table 3 comprising five distinct types of lexical cohesion and four distance types has a number of advantages. One is that it allows us to assess the cohesive force arising from each category independently. Two other assets are the non-categorial nature of the values, which enables us to see each type of lexical cohesion in terms of gradience between two relatively clearly defined poles, as well as the dissociation of *collocation* and *associative cohesion*, which in our view represent two different notions that are often coalesced. In addition, the taxonomy seems easier to apply. Not only have two categories (superordinates and general nouns) been subsumed under the general label *inclusion*, but the creation of different (sub)categories depending on whether there is identity or non-identity of reference has also been avoided. The reason is that this difference is not felt to influence whether a semantic relation does or does not exist in the discourse, although it may play a role in determining the strength of the relation.

LEXICAL DEVICES			TYPES OF TIES IN TERMS OF DISTANCE
Repetition (R)	Exact (ER)		Immediate (IM) Immediate-mediated (IM-MD) Remote (RM) Remote-mediated(RM-MD)
	Inexact (IR)	Inflectional (IIR)	
		Derivational (DIR)	
Synonymy (S)	Near-Synonymy (NS)		
	Propositional-Synonymy (PS)		
Opposition (O)	Complementaries (CpO)		
	Antonyms (AO)		
	Directional Opposites (DO)		
Inclusion (I)	Generalization (GI)		
	Specification (SI)		
Associative Cohesion (AC)			

Table 3. A combined corpus and systemic-functional analysis of lexical cohesion

Section 3 describes and illustrates the categories presented in Table 3 as the results of this corpus-based study are unveiled. The data consisted of seven telephone conversations between intimates (TCs) (13,952 words) and seven multi-party broadcast discussions (BDs) (13,476 words), extracted from the private and public spoken categories respectively of the *International Corpus of English-Great Britain* (ICE-GB). In keeping the discourse-specific view adopted, each cohesive tie was manually assessed in relation to its context of occurrence. The possible interactions among the variables were statistically analyzed using (*Pearson*) *chi square tests* and *log linear analyses* as implemented by SPSS (Field 2009).

3. Discussion of findings

A total of 13,365 pairs of lexical cohesion occurring across (AT) or within (WT) the turns were identified and assigned to the fifteen (sub) types introduced in Section 2. The findings displayed in Tables 4 and 5 below firstly reveal that BDs, examples of multiparty conversations, have almost six times as many lexical cohesion ties as TCs, represent-

ing dyadic conversations (11,199 *vs.* 2,166, with a mean frequency of 0.8 *vs.* 0.15 respectively). This association was reported as statistically significant by the chi square values both within turns ($\chi^2(4) = 46.10$, p < .001) and across turns ($\chi^2(4) = 29.04$, p < .001).

These results seem to contradict the view that the number of speakers affects the location of pairs across or within turns (Tanskanen, 2006: 104), as well as the hypothesis that the more speakers, the fewer the ties of (lexical) cohesion (Kerbrat-Orecchioni, 2004: 6; Tanskanen, 2006: 95, 102, 169). Instead, the data apparently suggest that either there exists a direct proportion between these two variables, which in our sample is statistically confirmed only in the R category by the Pearson correlation coefficient (r = .829, N = 15, p < 0.05), or else it is the generic differences between BDs and TCs, specifically regarding tenor (relations among the participants) and mode (channel and rhetorical purpose), which could explain such skewed distribution. The relative scarcity of lexical cohesion devices in TCs can be due to their nature as spontaneously unplanned private dyadic online machine-mediated conversations that are held between either relatives or friends with a predominantly phatic purpose (e.g. to contact someone or to keep in touch with someone) without paralinguistic cues. Therefore, it would appear that in the TCs under inspection coherence is achieved with little lexical cohesion. In contrast, BDs represent a pre-planned genre involving multiparty face-to-face public speech with two to five co-present discussants, who address each other upon the chairperson's dictates, generally with an informative-persuasive-instructive purpose in order to have an intended effect on the audience or addressees. Discussants collaboratively create the discourse at issue with generally a more defined script in mind, thereby increasing the use of lexical cohesion ties and chains across/within speakers and turns.

Another important finding observed in Tables 4 and 5 is that, despite their generic differences, the BDs and TCs under inspection replicate exactly the same frequency ranks of lexical cohesion devices: R > AC > I > S > O. Repetition (R) emerges as a powerful source of topic continuity and rhetorical intensification (Norrick, 1987; Stokes, 2004; Tannen, 2007). As can be seen in the instances of derivational repetition (DIR) reproduced in excerpts (11) and (12) from two BDs, participants repeat their own words (*self-repetition*), as in *culture – cultural*, or echo someone else's words (*other-repetition*), as in *United Nations-UN*, generally across the turns, to develop or to re-introduce the same or a temporarily

	ER n	IIR n	DIR n	R Sub total n (%)	AC Sub total n (%)	GI n	SI n	I Sub total n (%)	NS n	PS n	S Sub total n (%)	CpO n	AO n	DO n	O Sub total n %	Total n (%)
WT n (%)	556	79	31	**666**	**190**	19	85	**104**	46	2	**48**	21	2	11	**34**	**1,042 (9.3)**
AT n (%)	3,993	982	979	**5,954**	**2,482**	159	662	**821**	522	3	**525**	338	1	36	**375**	**10,157 (90.7)**
Total n (%)	**4,549**	**1,061**	**1010**	**6,620 (59)**	**2,672 (24)**	**178**	**747**	**925 (8.2)**	**568**	**5**	**573 (5.1)**	**359**	**3**	**47**	**409 (3.7)**	**11,199 (100)**

Table 4. Ties of lexical cohesion in BDs (raw frequencies)

	ER n	IIR n	DIR n	R Sub total n (%)	AC Sub total n (%)	GI n	SI n	I Sub total n (%)	NS n	PS n	S Sub total n (%)	CpO n	AO n	DO n	O Sub total n (%)	Total n (%)
WT n (%)	58	22	12	**92**	**64**	14	33	**47**	13	2	**15**	2	2	5	**9**	**227 (10.4)**
AT n (%)	666	337	41	**1044**	**576**	48	125	**173**	72	13	**85**	12	11	38	**61**	**1,939 (89.6)**
Total n (%)	**724**	**359**	**53**	**1,136 (52.5)**	**640 (29.6)**	**62**	**158**	**220 (10.1)**	**85**	**15**	**100 (4.6)**	**14**	**13**	**43**	**70 (3.2)**	**2,166 (100)**

Table 5. Ties of lexical cohesion in TCs (raw frequencies)

abandoned topic in order to provide an expansion or an evaluation of what is at issue. Repeating a local topic lessens the need for a global topic, facilitates micro-level information processing and emphasizes the repeated item(s). Closely related to this, another salient use of repetition is to agree/confirm and/or deny/reject an idea that has been grounded in discourse (Ochs Keenan, 1977; Norrick, 1987: 245-246). This usage is illustrated in (13), an extract of a TC in which B negates A's questions and asks for a clarification of the location of *Weymouth*, whereby a local topic is collaboratively created.

(11)   1: A: (2) To outsiders London seems one of the most vibrant <u>cultural</u>
        capitals of the world <,> a city (…) (4) The arts <u>culture</u> and entertain-
        ment industries together make up London's third largest employer after
        the Government (…).
(12)   6: C: (17) Uhm the hostilities end <,> when Saddam hu Hussein accepts
        <u>United Nations</u> resolutions (…).
        12: D: (43) It would be compatible with the <u>UN resolutions</u>.
(13)   13: A: (42) Did I tell you that I was <,> going for a job in in <u>Weymouth</u>.
        14: B: (43) No (44) Where is <u>Weymouth</u>.

Second in frequency are the ties of associative cohesion, which are here distinguished from collocation. *Collocates* or *lexical bundles* (Biber *et al.*, 1999) are pairs or strings of word forms that tend to co-occur with greater than random probability in their textual context, typically from four to six adjacent items (Stubbs, 2001; Stokes, 2004), although some authors also recognise them across textual boundaries (Scott and Tribble, 2006). The psychological processes involved in the selection of such co-occurrence patterns are explored in the theory of *lexical priming* (Hoey, 1991, 2001, 2005).

Associative cohesion (AC), on the other hand, operates across and within utterances and turns, as well as between longer stretches of discourse, through the *inferential processes* involved in the (re)construction of *triggers* and *frames* (Hawkins, 1978; Jordan, 1992; Fillmore and Baker, 2001). Triggers or prompts are lexical items that are explicitly materialized in texts (e.g. first-mention definite descriptions). Often repeating previous topics and based on the speakers' and hearers' shared knowledge, they make possible a set of associations of the generic relationship established between the trigger and its associate(s) or frames. The implication is that, while all collocates involve some kind of as-

sociative relation, not all associates need to be collocates. For not only can associates occur across longer spans, but also they may not always co-occur because the terms that are in associative relation in one text (e.g. in literary or metaphoric language) may not be associates in other text types or genres.

Our data suggest that AC is mainly employed to realize topic drifts and/or shifts. Excerpt (14) from a TC is a case in point, where the trigger *the theatre,* introduced by A and repeated by B evokes the frame of play writing and storytelling. This frame leads to a topic drift and becomes the local topic over B's turn 12 by forming a cohesive chain with lexical and grammatical nodes (underlined).

(14) Turn 11: Speaker A (male): (38) I don't know much about <u>the theatre</u>.
Turn 12: Speaker B (female): (39) <u>The theatre</u> (40) Right (41) Well I'm I'm writing <u>a a play</u> for them. (42) So uhm <,> so I 've kind of <,> been doing that (43) I did I did a bit of <u>storytelling</u> for uhm a shadow <u>puppet theatre</u> <u>company</u> from September to December <u>which</u> was really nice <,,> (44) And uhm <,> <u>that</u>'s all I've been doing really <,,> (45) Been pretty busy <,,> (46) How about you

Turning to Inclusion (I), the third most frequent lexical cohesion device in our sample, it also tends to be used to perform topic shifts and/or topic drifts. I comprises two complementary kinds of inclusive relations: *generalization* (GI), or specific-to-general relations, and of *specification*, or general-to-*specific* relations (McCarthy 1988). Transcript (15) from a TC shows that both generalization (*London-places*) and specification (*places-Colchester, Essex Suffolk Weymouth, Weymouth* (co-specific or co-hyponymic *items*)) contribute to shifting attention to different places that A intends or is expected to visit.

(15) 70: B: (190) And so you're not you're not coming to <u>London</u> at all.
71: A: (193) But <,,> uhm <,,> the thing is I was hoping to sort of <,,> sort of <,> visit some <u>places</u> around here and also some some people who uhm
72: B: (194) Uhm (195) What (196) <u>Colchester</u>
73: A: (197) Yeah (198) Well uh more than that <u>Essex Suffolk</u> uh
74: B: (199) But but you're not coming up much from <u>Weymouth</u>
75: A: (200) Oh no no no n (201) What what (202) sorry I thought
76: B: (203) I mean you're going to stay all seven weeks in <u>Weymouth</u>
77: A: (204) Oh yeah yeah (205) I'll have to yeah yeah

Coming fourth in the frequency rank, synonymy (S) is conceived as a scalar notion referring to two different kinds of similarity of meaning: *near* and *propositional* or *attitudinal* (Lyons, 1977, 1981; Martin, 1992; Cruse, 2004). Under the heading near-synonymy (NS) fall items that are situated along a graduated scale or otherwise occupy different positions with regard to a prototype center; whereas propositional synonymy (PS) refers to synonyms that differ in expressive meaning or stylistic level. Thus, the lexical tie *musician-composer* in BD (16) exemplifies NS because discussants C and B are interpreted as using these two terms as synonymous even if their prototypes profile different features as salient, namely to compose a piece of music and to play an instrument, respectively. On the other hand, the lexical tie *burst into tears-cry* instantiates PS. Two different verbal expressions are used by the same speaker in this TC over immediate moves of the same turn to avoid redundancy, and by so doing the same scene is pictured differently: *burst into tears* construes a more dramatic and emotional situation, while *cry* describes it more objectively. Furthermore, both terms participate in a cohesive chain (underlined) with grammatical and lexical nodes collaboratively created by A and B to develop the local topic, namely that *she* needs to keep her temper in front of her mother. As can be seen from the two examples provided, S is mostly used as a topic continuity strategy in the two conversational genres under analysis.

(16)   5: C: (6) (…) but I can't imagine that one composer would be employed for the entire project.
29: B: (95) I think the the secret for a director if he wants to talk to a musician is is actually putting the right penny in the right slot and and making a suggestion (96)

(17)   33: A: (110) And uh then when I asked her she burst into tears <,,> (111) And said look you're not to cry <,> (112) I said in front of your mother don't you (113) Not to ever break down because if she 's going to fight it
32: B: (114) No (115) Ah (116) Oh (117) Uhm (118) No (119) She's got to have (120) Yes yes
33: A: (121) Yes (122) She's got to have strength (123) So we all told her that but <,> (124) And I'm frightened to ask her in case it you know (125) I'll wait (126) I'll bring it up when she's on her own

Let us now turn to *opposition* (O), the least frequent lexical cohesion device in our corpus. O comprises three main kinds of semantic relations of exclusion*: complementaries, antonyms* and *directionals* (Cruse 2004). Complementaries (CpO) are absolute opposites in which the denial of one implies the admission of the other, as in the case of *women-men* in BD (18). Alternatively, antonymy (AO) refers to opposites which are fully gradable and incompatible, such the adjectives that create the lexical chain *difficult-difficult-easy* in TC (19), while directionals (DO) either evoke opposite directions (*reversives*), like *downturn-upturn* in (20), or denote reversed relationships (*converses*), as in the references denoted in *recruiting-redundant* in (21).

(18)   34: D: (148) I can remember when common sense said that for instance women were weaker than men (149) women shouldn't wear trousers.

(19)   36: C: (121) I I find it quite difficult to put into words (…) (132) And it's very difficult to do (133) (…) when you've got the sort of Ruritanian costumes that we we we adopted it's very easy.

(20)   8: B: (24) (…) the downturn in business was still expected to continue into the autumn (…).
        18: C: (47) And I think it's also dangerous as well to assume at the moment that the upturn in the economy will run parallel with uh employment levels <,>.

(21)   18: C: (49) (…) the the squeeze on profitability companies are no longer recruiting (50) in fact they are actually making people redundant.

The illustrations offered here aim to portray O as a topic continuity strategy. Its scarcity is probably due to the constraints on online communication, as well as to the fairly informal register of both BDs and TCs examined, which disfavour the usage of O or S, two relatively sophisticated and more demanding topic continuity techniques in processing terms than repetition.

Lastly, considering distance, Tables 4 and 5 above also reveal that most lexical cohesion ties occur across turns in both BDs and TCs. A more fine-grained quantitative analysis of this variable is offered in Table 6, which, with the exception of IM links in BDs, confirms the tendency for all the other distance ties to show more links across than within turns. This association was proved statistically significant by the results of three-way log linear analysis and Pearson chi square tests (BD $\chi^2(4) = 24.90$, $p < .001$; TC $\chi^2(4) = 37.40$, $p < .001$).

Distance ties	BDs			TCs			
	WT n	AT n	BD Subtotal n (%)	WT n	AT n	TC Subtotal n (%)	Total n (%)
RM-MD	397	8,760	**9,157 (81.8)**	62	1495	**1,557 (71.9)**	**10,714 (80)**
RM	271	666	**937 (8.3)**	59	291	**350 (16.2)**	**1,127 (9)**
IM	280	261	**541 (4.8)**	93	97	**190 (8.8)**	**731 (6)**
IM-MD	94	470	**564 (5)**	10	57	**67 (3.1)**	**631 (5)**
Subtotal n (%)	**1,042 (9.3)**	**10,157 (90.7)**	**11,199 (100)**	**240 (11.1)**	**1,926 (88.9)**	**2,166 (100)**	**13,365 (100)**

Table 6. Distance lexical cohesion ties across BDs and TCs

It can also be seen that the remote-mediated variant is the most common distance type across BDs (9,157 (81.8%)) and TCs (1,557 (71.9%)). Second in frequency come remote ties (937 (8.3%) in BDs vs. 350 (16.2%) in TCs), followed by immediate-mediated (564 (5%) in BDs vs. 190 (8.8%) in TCs) and immediate links (541 (4.8%) in BDs vs. 67 (3.1%) in TCs). Extracts (22) and (23) below represent the two categories newly posited in this model, that is, *immediate-mediated* and *remote-mediated* ties, to account for the difference between cohesive links occurring in consecutive (*like* (turn 94)-*like* (turn 95)) and non-consecutive or distant moves (*October* (turn 16)-*September* (turn 63)) that are mediated by one or several ties (*liked* (turn 94) in 22), and the double repetition of *months* (turn 60) in 23), respectively.

(22)   33: A: (94) (…) have you thought that perhaps he <,> won't like the new you you know perhaps that he liked the fact that you were so dependent on him (95) Oh no he'll like me.

(23)   4: B: (21) And she died the following October and then I got interested in it.

16: B: (60) And then for <u>months</u> and <u>months</u> it went like that (….) (63) And when I started again in the <u>September</u> I said now if nothing comes you've had it chum.

These findings deviate from those found in more stereotypical forms of dialogue, such as informal dyadic face-to-face conversations, which were found to favor reiteration pairs within the turns (Tanskanen, 2006: 104). The numbers reported here rather emphasize the highly collaborative nature of the TCs and BDs examined, as lexical cohesion is primarily constructed by different speakers to bridge the gaps generally between long spans of conversations. As a corollary, it could perhaps be postulated that connectivity within the turns in BDs and TCs is generally performed through grammatical cohesion, but this division of labor between the two types of cohesion awaits further investigation.

4. Conclusions

A five-fold taxonomy of lexical cohesion (repetition, synonymy, opposition, inclusion and associative cohesion) recognizing four kinds of distance ties (immediate, remote, immediate-mediated and remote-mediated) has been attested in a corpus consisting of broadcast discussions and telephone conversations.

The results confirm that lexical cohesion is affected by generic differences since broadcast discussions have six times as many instances of lexical cohesion as telephone conversations. However, the two genres display roughly symmetrical frequency patterns for the five categories and distance types of lexical cohesion scrutinized, which seems to profile a tendency in the spoken mode that transcends cross-generic differences. Repetition emerges as the most recurrent lexical cohesion strategy followed by associative cohesion, synonymy, inclusive and opposition relations, and all of these lexical cohesion devices are largely produced across the turns in remote(-mediated) ties. Furthermore, it has been explained that, while repetition is profiled as the predominant topic continuity device together with synonymy and opposition, associative cohesion and inclusion mostly lead to topic shifts and/or topic drifts across the two genres.

In sum, the chapter has integrated quantitative and qualitative evidence to demonstrate the effectiveness of the approach and, more

generally, the relevance of (lexical) cohesion in coherent communication. Further research is necessary to test the framework on more solid grounds in different discourse settings. Such investigations will not only prove the benefits of the model proposed so far, but will also shed new light on the interaction between (lexical) cohesion, coherence and genre. It is hoped researchers would find this approach stimulating and continue optimizing the theory of cohesion, adding –as Hoey (1991: 245) puts it– "what they have to say to the endlessly expanding net that is the sum of all human discourse."

Acknowledgements

References

Baker, M. 1992. *In Other Words*. London: Routledge.
Biber, D.; Johansson, S.; Leech, G.; Conrad, S. & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
Carter-Thomas, S. 2008. Teaching coherence through genre. *De la France au Quebec - l'Ecriture dans tous ces etats*. IUFM Poitou-Charentes (12-15 November), 1-11.
Cruse, D. A. 2004. *Meaning in Language: An Introduction to Semantics and Pragmatic*s (2nd ed.). London / Oxford: University Press.
Dontcheva-Navratilova, O. & Povolná, R. 2009. *Coherence and Cohesion in Spoken and Written Discourse.* Cambridge: Scholars Press.
Downing, A. 2000. Talking topically. In Downing, A.; Guijarro, J. M. & Albentosa, J. I. (ed.) *Talk and Text. Studies on Spoken and Written Discourse*. Cuenca: University of Castilla-La Mancha, 31-50.
Ebrahimpourtaher, A. & Eissaei, S. 2013. Awareness of lexical cohesive devices in text and reading comprehension. *International Journal of Educational Research and Technology* 4(2), 63-65.
Field, A. 2009. *Discovering Statistics Using SPSS* (3rd ed.). Los Angeles / London / New Delhi / Singapore / Washington: Sage.

Fillmore, Ch. J. & Baker, C. F. 2001. Frame semantics for text understanding. In *Proceedings of WordNet and Other Lexical Resources Workshop*, NAACL. Pittsburgh, June. Gzipped Postscript (112 KB) PDF (189 KB)

Flowerdew, J. 2010. Use of signalling nouns across L1 and L2 writer corpora. *International Journal of Corpus Linguistics* 15(1): 36-55.

Flowerdew, J. & Mahlberg, M. (ed.). 2009. *Lexical Cohesion and Corpus Linguistics*. Amsterdam: John Benjamins.

Fraser, B. 2005. *Towards a Theory of Discourse Markers*. Retrieved from http://people.bu.edu/bfraser/

Francis, G. 1986. *Anaphoric Nouns* (Discourse Analysis Monograph 11). English Language Research, Birmingham: University of Birmingham.

Gómez González, M. Á. 2010. Evaluating lexical cohesion in telephone conversations. *Discourse Studies* 12(5): 599-624.

Gómez González, M. Á. 2011. Lexical cohesion in multiparty conversations. *Language Sciences* 33(1): 167-179.

Gómez González, M. Á. 2013. A Reappraisal of Lexical Cohesion in English Conversations. *Applied Linguistics* 34(2): 128-150.

Gutwinski, W. 1976. *Cohesion in Literary Texts: A Study of Some Grammatical and Lexical Features of English Discourse*. The Hague / Paris: Mouton.

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

Halliday, M. A. K. & Hasan, R. 1985. *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective*. Oxford: Oxford University Press.

Halliday, M. A. K. 1985. *An Introduction to Functional Grammar*. London: Edward Arnold.

Halliday, M. A. K. 1994. *An Introduction to Functional Grammar* (2nd ed.). London: Edward Arnold.

Halliday, M. A. K. & Matthiessen, M. I. M. 2014. *An Introduction to Functional Grammar* (4th ed.). London: Edward Arnold.

Hawkins, J. A. 1978. *Definiteness and Indefiniteness: A Study in Reference and Grammaticality Prediction*. London: Croom Helm.

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

Hoey, M. 2001. *Textual Interaction. An Introduction to Written Discourse Analysis*. London: Routledge.

Hoey, M. 2005. *Lexical Priming. A New Theory of Words and Language*. London: Routledge.

Jordan, M. P. 1992. An integrated three-pronged analysis of a fund-raising letter. In Mann W. C. & Thompson, S. A. (ed.) *Discourse Description. Diverse Linguistic Analyses of a Fund-Raising Text*. Amsterdam: Benjamins, 171-227.

Kerbrat-Orecchioni, C. 2004. Introducing polylogue. *Journal of Pragmatics* 36(1): 1-24.

Kunz, K. 2015. Cohesion in English and German. A Corpus-Based Approach to Language Contrast, Register Variation and Translation. Habilitationsschrift. Universität des Saarlandes.

Lewis, M. 1993. *The Lexical Approach. The State of ELT and a Way Forward*. Hove: Language.

Lyons, J. 1977. *Semantics* 1. Cambridge: University Press.

Lyons, J. 1981. *Language, Meaning and Context*. London: Fontana.

Martin, J. 2001. Cohesion and Texture. In Schiffrin, D.; Tannen, D. & Hamilton, H. E. (ed.) *The Handbook of Discourse Analysis*. Malden, MA.: Blackwell Publishers, 35-53.

Martin, J. R. 1992. *English Text: System and Structure*. Philadelphia / Amsterdam: John Benjamins Publishing Company.

McCarthy, M. 1988. Some vocabulary patterns in conversation. In Carter, R. & McCarthy, M. (ed.) *Vocabulary and Language Teaching*. London / New York: Routledge, 197-208.

McNamara, D.; Louwerse C. & Graesser, A. 2002. Coh-Metrix (Version 2.0) [Software]. Memphis, TN: University of Memphis, Institute for Intelligent Systems. Available from http://cohmetrix.memphis.edu/cohmetrixpr/index.html

Moreno, A. 2003. The role of cohesive devices as textual constraints on relevance: A discourse-as-process view. *International Journal of English Studies* 3(1): 111-165.

Morris, J.; Beghtol, C. & Hirst, G. 2003. Term relationships and their contribution to text semantics and information literacy through lexical cohesion. In *Proceedings of the 31st Annual Conference of the Canadian Association for Information Science*. Halifax, Nova Scotia, June 1-June 4, 2003, 153-168.

Norrick, N. R. 1987. Functions of repetition in conversation. *Text* 7: 245-264.

Ochs Keenan, E. 1977. Making it last: repetition in children's discourse. In Ervin-Tripp, S. & Mitchell-Kernan, C. (ed.) *Child Discourse*. New York: Academic Press, 26-39.

Sanders, T. & Pander Maat, H. 2006. *Cohesion and Coherence: Linguistic Approaches. Encyclopedia of Language and Linguistics* (2nd ed.). Elsevier: London.

Scott, M. & Tribble, C. 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Philadelphia: John Benjamins.

Schmid, H.-J. 2000. *English Abstract Nouns as Conceptual Shells: From Corpus to Cognition*. Berlin: Mouton de Gruyter.

Steiner, E. 2015. Contrastive studies of cohesion and their impact on our knowledge of translation. In Zhang, M. & Jeremy, M. (ed.) *Discourse Analysis in Translation Studies*. Special issue of *Target* 27(3). International Journal of Translation Studies. Amsterdam: John Benjamins.

Stokes, N. 2004. *Applications of Lexical Cohesion Analysis in the Topic Detection and Tracking Domain*. Ph.D. dissertation, Department of Computer Science, University College Dublin, Dublin.

Stenström, A. B. 1994. *An Introduction to Spoken Interaction*. London: Longman.

Struthers, L.; Lapadat, J. & MacMillan, P. 2013. Assessing cohesion in children's writing: development of a checklist. *Assessing Writing* 18(3): 187-201.

Stubbs, M. 2001. Computer-assisted text and corpus analysis: Lexical cohesion and communicative competence. In Schiffrin, D. *et al.* (eds) *Handbook of Discourse Analysis*. Oxford: Blackwell, 304-20.

Taboada, M. 2004. *Building Coherence and Cohesion: Task-Oriented Dialogue in English and Spanish.* Amsterdam / Philadelphia: John Benjamins.

Tadros, A. 1985. *Prediction in Text*. (Discourse Analysis Monograph 10). English Language Research, Birmingham: University of Birmingham.

Tannen, D. 2007. *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse* (2nd ed.). Studies in Interactional Sociolinguistics, 6. Cambridge [England]: Cambridge University Press.

Tanskanen, S.-K. 2006. *Collaborating towards Coherence: Lexical Cohesion in English Discourse.* Amsterdam / Philadelphia: John Benjamins.

Winter, E. O. 1974. Replacement as a function of repetition: a study of some of its principal features in the clause relations of contemporary English. Unpublished PhD thesis, University of London.

Winter, E. O. 1977. A clause relational approach to English texts: a study of some predictive lexical terms in written discourse. *Instructional Science* 6: 1-92.

Winter, E. O. 1979. Replacement as a fundamental function of the sentence in context. *Forum Linguisticum* 4(2): 95-133.

Yankova, D. 2006. Semantic relations in statutory texts: A study of English and Bulgarian. *SKY Journal of Linguistics* 19: 189-222.

Xi, Y. 2010. Cohesion studies in the past 30 years: development, application and chaos. *Language, Society and Culture* 31(1): 139-147.