International Conference on Corpus Linguistics (CILC2013)

Both On and Under the Surface of Discourse: Tagged Corpora for the Functional Description of Conjunctive Language

Andy Cresswell[1]

University of Bologna, Department of Interpreting and Translation, 47100 Forlì, Italy

Abstract

This paper reports an eight-stage procedure for tagging small specialised corpora with logical relations, grounded in the coding of two corpora of argumentative writing – one learner corpus, and one expert corpus. The tagset was developed out of Rhetorical Structure Theory. Tagging involved adjusting mark-up added by the *RST Tool* tree diagram program to produce a corpus with cocoa tags easily searchable with KWIC concordancing software. Results show disambiguation of form-function relationships in conjunction and a wide range of types of conjunctive language. Potential further applications in research and teaching are discussed.

Keywords: learner corpora; EAP; logical relations; conjunction; rhetorical structure

1. Introduction: early and late stage analysis of conjunctive language

Conjunctive language is the surface realisation of underlying discourse relations – it functions as a message in which writers tell readers the type of logical relation that should be understood between two statements. Conjunctive language has been the focus of a number of corpus studies (e.g. Altenberg & Tapper, 1998; Bolton et al, 2002). These studies have all involved late-stage analysis, in which exponents like *however*, already theorised as having a conjunctive language function, are counted up and compared between various corpora. Researchers have used these counts to find that there is general overuse by second language learners of conjunctive connectors (Bolton et al., 2002:165), or have concluded that there is learner underuse of certain semantic types – for example, of contrastive and resultive connectors (Altenberg & Tapper, 1998:91).Late-stage analysis can be contrasted with early-stage analysis – what Sinclair (2001: xi) terms "early human intervention". In my early stage approach, logical relations are classified and tagged first and present a ready discourse framework for subsequent corpus linguistic analysis. This approach, combining function and form, is advantageous compared with late-stage analysis, which has to imply function on the basis of form – particularly since forms are polypragmatic. *In fact* is a case in point, being used both in evidence and replacive (or correction). Additionally, in early-stage analysis the lack of pretheorisation of what functions forms have allows the linguist to identify conjunctive roles in a wide range of language in addition to the conventional categories of conjuncts (like *however*) and subordinators (like *although*).

[1]* Corresponding author. *E-mail address:* cresswel@sslmit.unibo.it Selection and peer-review under responsibility of CILC2013.

2. Research aims and framework

My research aims were to use logical relation tagging to separate out the functions of polypragmatic forms, to reveal the full range of language used for conjunctive purposes, and to examine expert-learner differences. In the absence of corpora already tagged for logical relations, I constructed two small, specialised corpora (as theorised by Flowerdew, 2004) and tagged them for logical relations myself. One corpus was CRANE (corpus of non-empirical research articles), and the other was ALCASE (advanced learner corpus of argumentative student essays). Both contained 65000 words and shared the source-based argumentative discourse type.

Because logical relations are discourse phenomena, I sought a discourse oriented framework for coding the data, in terms both of pragmatic theory and of research processes. In terms of theory, the right framework was provided by Rhetorical Structure Theory, or RST (Taboada & Mann, 2006a). RST accounts for how texts serve authors' purposes, provides rigorous pragmatic definitions (see example in Appendix A), and potentially covers all logical relations in texts (Mann et al., 1992:41). In terms of research processes, coding was carried out by constructing tree diagrams, which enabled the researcher to visualise the relations analysed. Moving from raw texts through tree diagrams to texts with logical relation tags involved eight processes. These are described in the next section.

3. The 8 part tagging procedure

Although the eight processes of the tagging procedure are described as eight numbered steps, it should be understood that the sequence is chronological only in an approximate sense. Because of the integral role that was played by evaluation in the tagging procedure, the shift by the researchers from one process to another was recursive, and thus not necessarily in strict numerical order. It is also worth noting that processes 1 to 7 are essentially text by text, while procedure 8 involves processing the whole corpus.

3.1 Insert raw texts in a tree diagram program

The program selected to code CRANE and ALCASE was *the RST Tool* (O' Donnell, 2000), version 3.45. This program permits the importation of corpora in the form of text files, as well as of metafiles containing lists of relation names. *The RST Tool* converts the imported text files into XML files, in which tree diagram structures and tags showing logical relations are coded as attributes. The XML format of *the RST Tool* permits visualisation of the coherence structure of an entire text. This means that logical relations can be perceived even when connecting two relatively distant text parts. Use of tree diagrams arguably has the advantage of reducing subjectivity in the tagging process – because the trees provide a visual reference of text structure, data coders do not have to rely entirely on memory when tracing and naming logical relations. In this way coding decisions gain in reliability because they are based on observation - that is, the observed relation of any text span with any other text span, or with the whole text.

3.2 Divide texts into discourse units

Most conjunctive language refers to relations between units at clause level and above, so the clause was chosen as the basic discourse unit. This was also convenient, since the RST Tool segments texts on the basis of punctuation, which usually coincides with clause boundaries. Since the research aims included identifying all possible conjunctive language, some exceptions were made, to include relations like exemplification which often involve noun phrases. The exceptions were listed in protocols (consultable on line at http://amsacta.unibo.it/3654/).

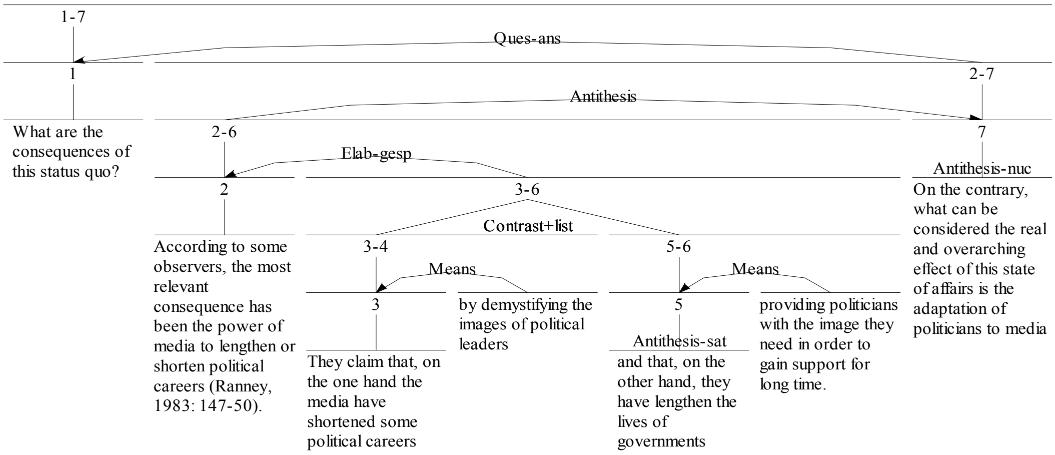
3.3 Train analysts

I and another analyst used a basic set of RST relation definitions, largely drawn from *the RST website* (Mann & Taboada 2005, 2012), to independently code texts corresponding to approximately the first 5% of words of both corpora as tree diagrams. We then compared and reached agreement. This enabled us to develop a shared understanding of RST technicalities like nuclearity, parataxis, hypotaxis and the way that definitions are couched in terms of writers' intended effects on readers.

3.4 Test the functional paradigm against the discourse

RST relations are hyperonymous (Mann et al. 1992:46) – they may be subdivided to match particular logical relations occuring in specific discourse types. To ensure we achieved a complete set of relations suitable for argumentative texts, we compared analyses on the first 15% of words in both corpora and, when necessary, improved the discreteness of the relation definitions. This involved consulting research on logical relations, such as Carlson and Marcu (2001), Crombie (1985), and Martin (1992), then writing supplementary relation definitions in the rigorous RST style, specifying nuclearity and the writer's intended effects on the reader. The complete paradigm of relations suitable for argumentative texts, used in the final logical relations tagset, is shown in Appendix B.

Fig. 1: How RST diagrams are constructed



1-7
Ques-ans
1
2-7
Antithesis
What are the consequences of this status quo?
2-6
Elab-gesp
7
Antithesis-nuc
2
3-6
On the contrary, what can be considered the real and overarching effect of this state of affairs is the adaptation of politicians to media
Contrast+list
According to some observers, the most relevant consequence has been the power of media to lengthen or shorten political careers (Ranney, 1983: 147-50).
3-4
Means
5-6
Means
3
by demystifying the images of political leaders
5
providing politicians with the image they need in order to gain support for long time.
They claim that, on the one hand the media have shortened some political careers
Antithesis-sat
and that, on the other hand, they have lengthen the lives of governments

3.5 Build tree diagrams

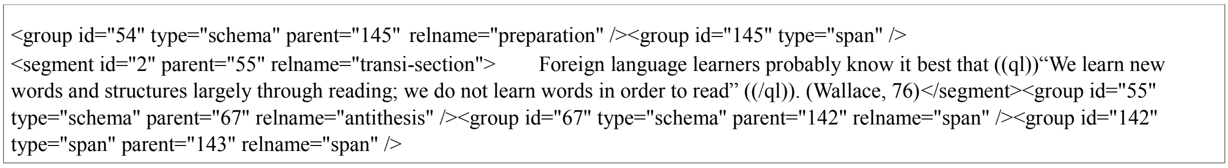
An extract from an ALCASE tree diagram is shown in Fig. 1. Fig.1 shows how tree diagrams are built up hierarchically, with height in the diagram reflecting the relative prominence in overall text coherence of given groups of discourse units (known as spans – see for example 2-6 in Fig. 1). Logical relations are drawn by clicking on the discourse units concerned and selecting a relation name from drop-down lists of hypotactic or paratactic relations (in Fig. 1, the relation between 3-4 and 5-6 is paratactic, and that between 2-6 and 7 is hypotactic). Relation names come from metafiles which are put into the program by the researcher.

3.6 Adapt to the software

The *RST Tool* was flexible enough to handle some aspects of logical linking it was not conceived for. For example, first, some pairs of discourse units were linked by two or more relations; the solution was to type in multiple relation tags (see 3-6 in Fig.1). Second, as a reading of Martin (1992:263-4) predicts, there were sometimes contingent lateral links between units in different main branches of the tree. This problem was solved by using the "schema element" facility to tag the separate discourse units both with the appropriate relation name and with a numbered reference to the other linked unit. With such adaptations, all the logical relations perceived were actually tagged.

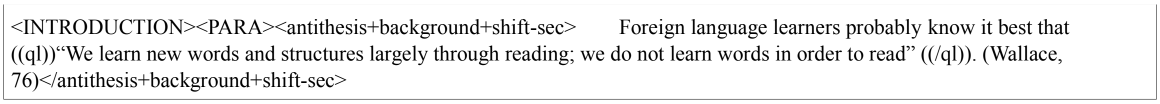
3.7 Check for consistency

Because of RST's hierarchical nature, classifications depend on previous decisions. The normal applied linguistic method of independent analysis of a proportion of texts would therefore have multiplied divergences exponentially. So the system adopted was for a second analyst to rigorously check a proportion of the work of the first analyst (chosen by lot), with subsequent discussion and revision, reinforced by referral of problematic text sequences. Overall, approximately one third of CRANE and ALCASE was checked, with a more than adequate index of agreement of 0.98 (measured according to alterations on checking - see details in Appendix C).



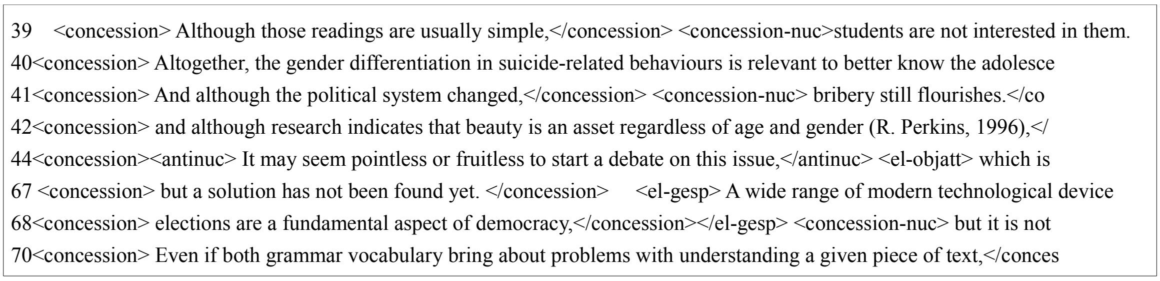
```
<group id="54" type="schema" parent="145"  relname="preparation" /><group id="145" type="span" />
<segment id="2" parent="55" relname="transi-section">       Foreign language learners probably know it best that ((ql))“We learn new words and structures largely through reading; we do not learn words in order to read” ((/ql)). (Wallace, 76)</segment><group id="55" type="schema" parent="67" relname="antithesis" /><group id="67" type="schema" parent="142" relname="span" /><group id="142" type="span" parent="143" relname="span" />
```

Fig. 2: Diagram attributes before automatic conversion



```
<INTRODUCTION><PARA><antithesis+background+shift-sec>       Foreign language learners probably know it best that ((ql))“We learn new words and structures largely through reading; we do not learn words in order to read” ((/ql)). (Wallace, 76)</antithesis+background+shift-sec>
```

Fig. 3: After automatic conversion: the tagged corpus



```
39    <concession> Although those readings are usually simple,</concession> <concession-nuc>students are not interested in them.
40<concession> Altogether, the gender differentiation in suicide-related behaviours is relevant to better know the adolesce
41<concession> And although the political system changed,</concession> <concession-nuc> bribery still flourishes.</co
42<concession> and although research indicates that beauty is an asset regardless of age and gender (R. Perkins, 1996),</
44<concession><antinuc> It may seem pointless or fruitless to start a debate on this issue,</antinuc> <el-objatt> which is
67 <concession> but a solution has not been found yet. </concession>     <el-gesp> A wide range of modern technological device
68<concession> elections are a fundamental aspect of democracy,</concession></el-gesp> <concession-nuc> but it is not
70<concession> Even if both grammar vocabulary bring about problems with understanding a given piece of text,</conces
```

Fig. 4: A KWIC concordance from ALCASE with relation tags and conjunctive language (extract)

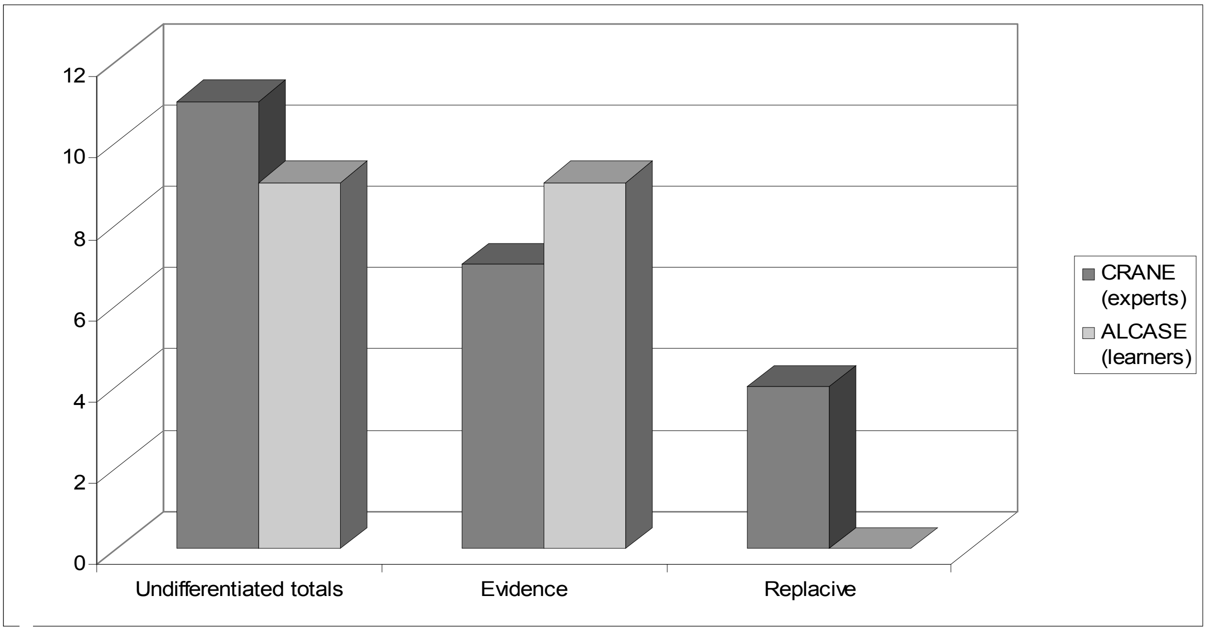


Fig. 5: *In fact* in CRANE and ALCASE: frequencies

3.8 Automatically convert

Although the logical relation tags are all visible at the outset in each diagram's *.rst* file, if it is read in a text editor, the problem is that the tags are dispersed between the text and the footer, and the quantity of non-logical-relation attributes (required for the diagram graphics) makes KWIC searching impractical (for an illustration, see Fig. 2). So the unwanted attributes are automatically removed, and the wanted tags aligned next to their matching discourse units, using Excel macros, Perl scripts and regular expressions, before a final manual check. Afterwards, what remains are the original texts with logical relation tags, as shown in Fig. 3. It is then possible to use KWIC concordancers such as *Concord* (Scott, 2011) to view logical the relation tags alongside matching conjunctive language (as shown in Fig.4).

4. Results: tagged corpus results versus late-stage results

A first result of the examination of KWIC concordances of the logical relation tags was to permit disambiguation of accounts of the use of polypragmatic logical connectors. This can be seen from Fig. 5, showing raw frequencies of *in fact* in CRANE and ALCASE. The left hand pair of bars shows frequencies undifferentiated for logical relation, as would be obtained in late-stage analysis; the centre and right hand pairs of bars show frequencies in concordances of the evidence and replacive tags. The undifferentiated frequencies suggest that there is little difference between expert (CRANE) and learner (ALCASE) use. The frequencies per relation, however, show that although learners and experts used *in fact* at similar frequencies in evidence, in replacive learners did not use it at all. In this way, concordances of corpora tagged for logical relations reveal information about learners' apparent lack of knowledge of one particular function of an item of conjunctive language, information that is masked in late-stage analysis.

A second result is to permit observation of the whole range of language used to signal logical relations. In late stage analysis, conjunctive forms are defined in advance – they are mainly conjuncts and subordinators, collectively referred to as sentence connectors. In relationally tagged corpora, examination of concordances with logical relation tags as search words reveals a whole range of conjunctive language, in addition to sentence connectors. This can be seen from Table 2, which shows the substantial percentages of statements of four logical relations in CRANE (the relations are defined in Appendices A and D) that contained what I loosely call "open class" signals, that is, conjunctive language that is neither subordinators, nor conjuncts, nor co-ordinating conjunctions, nor prepositions . The range of "open class" items can be grasped from Table 3, which shows ten different "open-class" categories found to be signalling antithesis in CRANE.

Table 2: "Open-class" conjunctive signals in CRANE

	Antithesis	Replacive	Evidence	Reinforcement
relation statements using open-class tokens (discrete single words + phrases), %	58	19	21	27

Table 3: Open-class language signalling antithesis

CONJUNCTIVE LANGUAGE CATEGORY/ STRING / STRUCTURE (in order of no. of expert texts in which they occurred, with main categories in capitals, and examples in italics)	**PERCENTAGE OF ANTITHESIS STATEMENTS IN *CRANE***
ADJECTIVE: COUNTER-FACTIVE: *Flawed, inaccurate, incorrect*	19
DISCOURSE VERB : COUNTER-FACTIVE: *Presupposes, ignore, misunderstand*	8
'ALTERNATIVE' ARGUMENT OR REFERENCE SIGNAL: NON-MODAL: *It is also, elsewhere, other, some*	8
DISCOURSE VERB: WITH NEGATION/RESTRICTION: *But this does not mean that*	7
NOUN: COUNTER-FACTIVE OR OPPOSITIONAL: *Disadvantage, misunderstanding*	7
DISCOURSE VERB: OPPOSITIONAL (OR PARAPHRASE): *Questioned, challenged*	6
QUANTIFIER WITH NEGATION/RESTRICTION: *no empirical support*	5
IRONY EXPRESSION: *ironically*	3
Case negated: *this is not the case*	3
VERB OF LACK *Needs, requires*	2

A third result is insight into conjunctive phraseology. The combination of syntagmatic and paradigmatic analysis of logical relation tags made possible by concordances (Tognini-Bonelli, 2004:18) permits the observation of conjunctive language patterns associated with the signalling of particular relations. To give one simple example from the concession relation, there is the pattern “Anticipatory *it* plus factive adjective”, exemplified by “It is true that (there are still many wars nowadays)” (from ALCASE).

Inter-corpus comparison of such patterns provides insights into conjunctive language variation. The anticipatory *it* pattern was used in 1.8% of experts' concession statements (in CRANE), while the learners in ALCASE used it in 3.4% of theirs. Since the pattern represents authorial distance from the content, it is probable that learners use the pattern more because on the one hand they are striving for objective stance to please their tutors, while on the other they are expressing their felt peripheral status in the academic community.

Conversely, the data show how lack of distancing, in the sense of confident authoritativeness, is also a factor in conjunctive language choice. The experts of CRANE used factive verbs, showing commitment to the truth value of the content (for example *accept*, *agree*, *take into account*), in 4.4% of concession statements, as against only 0.3% in ALCASE. This sharp difference between experts and learners again seems attributable to different degrees of felt academic centrality.

In sum, compared with late-stage analysis, the analysis of corpora tagged for logical relations is more precise functionally and more open-ended in terms of the conjunctive language it reveals. As the above examination of the results for anticipatory *it* and factive verbs shows, a wider range of language also makes it possible to discuss pragmatic factors governing conjunctive language choice.

5. Conclusions and further applications

To summarise and develop, the eight-part method for tagging for logical relations permits a clear view, through KWIC concordances, of form-function relations in conjunction, in texts from a single genre. This in turn permits the disambiguation of form-function relations. Because what is viewable is a whole line of language next to the relation tag, all the different types of language involved in signalling conjunction become observable, including syntacto-semantic patterns associated with particular relations. As logical relations are universal rather than being confined to single genres of texts, it is therefore in turn possible to use corpora tagged for logical relations to study variation in conjunctive language between two corpora each of a different genre – provided both are of the same discourse type. This means fair comparison can be made even between related expert and learner texts - as was the case of the argumentative RA's and essays of CRANE and ALCASE. This comparability facilitates the use of relationally-tagged corpora for learner needs analysis in the teaching of EAP. Form-function combinations observed in experts' argumentative texts, but used less or not at all in learners' texts (for example use of *in fact* in replacive), can be selected for instructive purposes.

There is much potential for further application of the method. The same logical relations tagset could be applied unaltered to the tagging of other genres of the argumentative discourse type in the social sciences, such as project proposals, or, with appropriate modifications to take note of the combination of empirical with argumentative content, to Ph.D theses and empirical RA's. With the two latter genres, part four, development of a functional paradigm (see 3.4 above), might involve the coining of a few new relation definitions. But because these genres are to a substantial extent argued, the tagset would be largely the same.

A further potential application is to use the tagset worked out for CRANE and ALCASE for the encoding of logical relations in corpora of argumentative genres written in languages other than English. Although RST has already been applied to make tree diagrams in at least eleven languages (Taboada & Mann, 2006b:572), few of the corpora involved appear to have been composed of argumentative texts. Concordances of particular logical relations generated from such corpora could provide a resource for contrastive analysis of logical connection, to investigate such issues as the relative prevalence in comparable genres in given pairs of languages of adverbial or non-adverbial logical links, the degree of authoritative or objective oriented conjunctive language, the extent of transferability across languages of conjunctive language patterns, or the comparative frequency of explicitly expressed logical links compared with those that are expressed implicitly. In this context work could be done to investigate how software that permits the alignment of sentences in parallel corpora - such as *YouAlign* -might permit the examination of the translation of the language of logical connection, for instructional purposes or quality control. One problem that comes to mind here is the varying degree of polypragmaticity between supposedly equivalent conjunctive devices in different languages. To give an example from translation between Italian and English, there is *invece*, which can be translated either as *instead* (signalling an internal relation of replacive) or *on the other hand* (signalling an external relation of simple contrast). This problem could either be investigated by examining parallel translated sentences, or by generating concordances of *invece* and *on the other hand* and comparing the relation tags.

In addition to these potential advantages, there are of course disadvantages. The recursive nature of the process of definition, analysis and checking involved in tagging with RST means that the tagged corpus is built up very slowly. Log data from the tagging of CRANE and ALCASE showed that the two analysts took 1250 hours to tag the 130,000 words of CRANE and ALCASE combined, or about 10 hours per 1000 words. This implies allowing long time spans for RST tagging projects. An additional practical factor is that the need to invest time in training

RST analysts makes it imperative that any projects involving tagging for logical relations with RST are planned so that the tagging process can be brought to completion without having to change the research team. The imperative of consistency requires frequent coordination – this, together with the time and labour involved, suggests that early stage tagging for logical relations using the methodology described in this article is only practical for small corpora of specialised texts. The size of corpus could probably be increased beyond the ±65000 words of those in the current research, if there is modification of the methodology so that there is a larger team of analysts and use of appropriate statistical analysis for checking consistency of analysis by such large teams, as with the extended use of the kappa coefficient in Carlson et al. (2003:108). But to tag a large corpus (upwards of a million words) would almost certainly be impossible in terms of consistency of the analysis – for large corpora demand automatic processing rather than early-stage human analysis. It is possible, however, to envisage that the accumulated results of several replicated studies done on a single genre using early-stage RST tagging could identify combinations of words, phrases, collocations and patterns associated with each logical relation, in the way that the language items listed in Table 3 are associated with antithesis. Assuming a solution can be found to the problem of polypragmatic signals – perhaps through the analysis of co-text - the results of such replicated studies could be used to design algorithms that could automatically tag large corpora for logical relations. Then once the groundwork of research into the characteristic language had been done using small specialised corpora, large corpora could be used for the sorts of comparison of genres or disciplinary discourses, within and across languages, which were proposed earlier in this section.

In sum, the methodology of early-stage tagging of small corpora for logical relations makes heavy demands in terms of time and labour. But the effort is worth it. Using RST for early-stage tagging of small corpora should, in the medium term, provide results to feed into automatic tagging of larger corpora for logical relations. In the short term, first, early stage tagging for logical relations with RST creates scope for the use of commonly available KWIC concordancing software in intensifying research into and increasing understanding of the language of conjunction in particular genres. Second, because the universality of the RST classification scheme facilitates comparison of conjunctive language across different but related genres, the method of tagging for logical relations has a clear pedagogical application in the sense of learner needs analysis. Third, the method could be applied to contrastive analysis in Translation Studies, since it provides the potential to compare conjunctive language in similar genres between different languages too.

Acknowledgements

Acknowledgements are due to the following publishers, who kindly gave permission for material from the following articles, to which they hold the copyright, to be reproduced in CRANE: first to Cambridge University Press, for Metz, D. 2002, 'The politics of population ageing', *Political Quarterly,* 73(3) pp. 321-327; and secondly to the National Association of Social Workers, for Hartman, A. (1991), ‘Toward a redefinition and contextualization of the abortion issue’. *Social Work,* 36(6), 467-468.

Some of the raw learner essay data making up ALCASE was also used in the PELCRA project (Leńko-Szymańska 2007:259), and now forms part of its learner English corpus. For information about PELCRA (Polish and English Corpora for Research and Applications), see http://ia.uni.lodz.pl/plec/.

Appendix A: Full definition of the 'evidence' relation

Relation Name (source of definition)	**Constraints on either S or N individually**	**Constraints on N + S (or N + N, or S+S)**	**Intention of Writer**	***Examples***
Evidence	on N: R might not believe N to a degree satisfactory to W; on S: R believes S or will find it credible	R's comprehending S increases R's belief of N ; a deduction in N is drawn on the basis of some observation in S (including both empirical, or logical grounds, or reason)	R's belief of N is increased	***<S>Health and social care costs … increase with age, </S>*** *<N> so an ageing population might naturally be thought to imply a bigger burden for individuals, families and society. </N> (Metz, 2002)*

Key
N nucleus **R** reader **S** satellite **W** writer **/N** nucleus end **/S** satellite end

Appendix B: Logical relations used for the tagset

Addition
Addition- emphasis
Analogy
Antithesis
Antithesis-Complexification
Background
Cause-Nonvolitional
Cause-volitional
Circumstance
Concession
Condition- hypothetical
Condition- Open
Condition- time
Contingency
Contrast
Contrast-specular
Disjunction
Disjunction – contrastive
Elaboration: equating
Elaboration: abstraction-instance
Elaboration: classification
Elaboration: contrast
Elaboration: definition
Elaboration: exception
Elaboration: naming
Elaboration: object-attribute
Elaboration: process-step
Elaboration: set-example
Elaboration: set-member
Elaboration: whole-part
Enablement
Evaluation (reported, negative)
Evaluation (reported, positive)
Evaluation (negative)
Evaluation (positive)
Evidence
Evidence + Irony
Interpretation
Interpretation (reported)
Joint
Justify
List
Means
Motivation
Non-result
Otherwise
Preparation
Presentational sequence
Problem-Solution, Solution-Problem
Proportion
Purpose
Question-answer
Question-rhetorical
Reinforcement
Replacive
Restatement
Restatement-emphatic
Result-non-volitional
Result-volitional
Sequence
Summary
Unconditional
Unless

Appendix C: Reliability

CHECKING BY SECOND ANALYST	CHANGES MADE PER 1000 UNITS
Segments with superficial errors	4
Generalisable rethinkings	1
Non-generalisable changes in relation tags	14
Changes in unit boundaries	4
Total changes	23

Appendix D: RST: a glossary, including some logical relations

ANTITHESIS. Antithesis, which is similar to the Hallidayan notion of 'adversative', can be briefly defined as 'because of an incompatibility that arises from the contrast between nucleus and satellite, the reader regards the nucleus positively and the satellite negatively'. **Example:** *<satellite> so an ageing population might naturally be thought to imply a bigger burden for individuals, families and society.</satellite>****<nucleus> However, it turns out that the main reason why average health care costs appear to rise with age is not that we need much more care on account of our advancing chronological age...*** *(article: Metz 2002)*

CONCESSION. Concession can be defined as 'the writer intends readers to acknowledge a potential or actual incompatibility between the nucleus and satellite, yet regards the situations in each as compatible enough for readers to accept the situation in the nucleus' - a logical relation often signalled by *although*.

Example: *<satellite>It can be inferred from the data that although the communism in Poland has been toppled, </satellite>* ***<nucleus> corruption not only stayed but it is also developing. </nucleus>*** *(essay 00010072)*

NUCLEUS. In a pair of related discourse units, the nucleus is the unit that is related to units higher in the text hierarchy.

REINFORCEMENT. The writer intends the reader to perceive the satellite as increasing belief in a preceding claim – as belief which has already been induced by a statement in the nucleus, the satellite creates a sort of cumulative effect. **Example:** *<nucleus>Such discourse on rights can become increasingly convoluted but does not indicate a way to deal with this crucial conflict in a highly diverse society. </nucleus> <satellite>* ***Further, rights arguments tend, in their absoluteness, individualism, and insularity, to be silent with respect to personal, civic, and collective responsibility.</satellite> (article Hartmann 1991)***

REPLACIVE. A short definition of replacive is 'nucleus and satellite are in contrast, with an incompatibility such that the writer intends the satellite to be rejected by the reader, and the nucleus to substitute it'. This often implies correction. **Example:** *<nucleus> Moreover, methodologists seem to do their best to bring the real world to the classroom, </nucleus>* ***<satellite> and not any obscure pieces of material that is no longer relevant. </satellite>*** *(essay 00010064)*

SATELLITE. In a pair of related discourse units, the satellite is either the unit that is related to units lower in the text hierarchy, or it is the lowest unit in an absolute sense.

References

Altenberg, B. & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners' written English. In S. Granger (Ed.), *Learner English on computer* (pp. 80-93). London: Longman,.

Anthony, L. (2011). *AntConc* (Version 3.2.2). Tokyo, Japan: Waseda University. http://www.antlab.sci.waseda.ac.jp/

Bolton, K., Nelson G., & Hung J. (2002). A Corpus-Based Study of Connectors in Student Writing: Research from the International Corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics*, 7 (2), 165-182.

Carlson, L., & Marcu, D. (2001). *Discourse Tagging Reference Manual*. On-line document. University of Southern California Information Sciences Institute. ftp://128.9.176.20/isi-pubs/tr-545.pdf Last accessed 15/12/12.

Carlson, L., Marcu, D., & Okurowski, M. E. (2003). Building a discourse tagged corpus in the framework of Rhetorical Structure Theory. In J. van Kuppevelt & R. Smith (Eds.), *Current and New Directions in Discourse and Dialogue* (pp. 85-112.). Berlin: Springer.

Crombie, W. (1985). *Discourse and Language Learning: a Relational Approach to Syllabus Design*. Oxford: Oxford University Press.

Flowerdew, L. (2004). The argument for using English specialized corpora to understand academic and professional language. In U. Connor & T. Upton (eds.), *Discourse in the Professions: Perspectives from Corpus Linguistics* (pp. 13-32). Amsterdam: Benjamins.

Leńko-Szymańska, A. (2007). Past progressive or simple past? The acquisition of progressive aspect by Polish advanced learners of English. In E. Hidalgo, L. Quereda & J. Santana (Eds.), *Corpora in the Foreign Language Classroom* (pp.255-266). Amsterdam: Rodopi.

Mann, W., Matthiessen, C. & Thompson, S. (1992). Rhetorical Structure Theory and Text Analysis. In W. Mann & S. Thompson (Eds.), *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text* (pp. 39-78). Amsterdam: Benjamins.

Mann, W. & Taboada. M. 2005, 2012). The RST Web Site.http://www.sfu.ca/rst/. Last accessed 15 December 2012.

Mann, W. C. & Thompson S. (1988). Rhetorical structure theory: Towards a functional theory of text organisation. *Text* 8(3), 243-281.

Martin, J. (1992). *English Text: System and Structure*. Amsterdam: John Benjamins.

O'Donnell, Michael. (2000). RSTTool 2.4: A markup tool for Rhetorical Structure Theory. *First International Conference on Natural Language Generation (INLG'2000)* (pp. 253-256). Mitzpe Ramon, Israel. http://www.wagsoft.com/software.html

Scott, M., (2011). WordSmith Tools version 6, Liverpool: Lexical Analysis Software. http://lexically.net/wordsmith/version6/index.html

Sinclair, J.McH. (2001a). Preface: Small Corpus Studies and ELT: Theory and Practice. In M. Ghadessy, A. Henry & R. Roseberry (Eds.), *Small Corpus Studies and ELT: Theory and Practice* (pp. vii-xv). Amsterdam: John Benjamins.

Taboada, M., & Mann, W. (2006a). Rhetorical Structure Theory: Looking Back and Moving Ahead. *Discourse Studies*, 8 (3), 423-459.

Taboada, M., & Mann, W. (2006b). Applications of Rhetorical Structure Theory. *Discourse Studies*, 8 (4): 567–588.

Tognini-Bonelli, E. (2004). Working with corpora. In C. Coffin, A. Hewings, & K. O'Halloran (Eds.), *Applying English Grammar* (pp. 11-24). London: Arnold, pp. 11–24.

YouAlign. 2013. Terminotix Inc. www.youalign.com. Last accessed 23/5/13.