Amsterdam University of Applied Sciences, Faculty of Education
Department of English and Education

TO CONNECT OR NOT TO CONNECT

Corpus-driven analytical and design research into written connectivity skills of advanced secondary-school students of English in the Netherlands

THESIS FOR MASTER OF EDUCATION IN ENGLISH
David Geneste, February 2012
Student:
David Geneste (student in MEd in English and teacher of English)
Javakade 694, 1019 SG Amsterdam
M 06-17556511
E gnt@st-maartenscollege.nl

Supervisor:
dr. Eline Raaphorst (lecturer in Educational Research)
Amsterdam University of Applied Sciences (Hogeschool van Amsterdam)
Faculty of Education, Department of English and Education
PO Box 1025, 1000 BA Amsterdam
E eraaphorst@cna.uva.nl

School:
Sint Maarten’s Secondary School (Sint Maartenscollege)
Aart van der Leeuwkade 14
2274 KX Voorburg
T 070-3867272
E info@st-maartenscollege.nl

School supervisor:
Leonieke Albreghs (teacher of English)
E alb@st-maartenscollege.nl
CHAPTER 1: INTRODUCTION
1.1 Research context 7
1.2 Research problem 9
1.3 Objectives, scope and organisation of research 10
1.4 Terms and abbreviations used in this study 10

CHAPTER 2: CONNECTIVITY, CORPUS LINGUISTICS AND TESOL 12
2.0 Introduction 12
2.1 Connectivity
   2.1.1 Definitions and research 12
   2.1.2 Linguistic properties and learner problems 13
      2.1.2.1 Semantic properties and learner problems 13
      2.1.2.2 Syntactic properties and learner problems 14
      2.1.2.3 Stylistic properties and learner problems 15
   2.1.3 Connectivity teaching: current practices and limitations 15
   2.1.4 New starting points and directions in connectivity teaching 16
2.2 Corpus linguistics
   2.2.1 Definitions and research 19
   2.2.2 Corpus linguistic and TESOL 19
2.3 A corpus-based approach to connectivity acquisition and teaching 21

CHAPTER 3: ANALYTICAL RESEARCH 23
3.0 Introduction and theory 23
   3.0.1 Research into connectivity usage and analysis 23
   3.0.2 The identification and labelling of connectives 24
   3.0.3 Inappropriateness in connectivity 25
      3.0.3.1 Underuse of connectives 25
3.0.3.2 Overuse of connectives 26
3.0.3.3 Misuse of connectives 27

3.1 Procedure 28
3.1.1 Objectives 28
3.1.2 Research questions and conceptual model 28
3.1.3 Research plan, conditions and characteristics 29
3.1.4 Research instruments 31

3.2 Implementation 33
3.2.1 Data collection 33
3.2.1 Results 34
3.2.1.1 The ten most frequently used connectives 35
3.2.1.2 Misuse of connectives in learner corpora 36
3.2.1.3 The most obstructive connectives and misuse 40

3.3 Evaluation 40
3.3.1 Conclusions 40
3.3.2 Recommendations 43

CHAPTER 4: DESIGN RESEARCH 44

4.0 Introduction 44
4.0.1 Link with analytical research 44

4.1 Procedure 45
4.1.1 Hypothesis, research questions and conceptual model 45
4.1.2 Research plan 46
4.1.2.1 Types of research and triangulation 46
4.1.2.2 Proceeding of the research 46
4.1.3 Research instruments 49
4.1.4 Description of intervention 49
4.1.4.1 Methodological starting points and framework 49
4.1.4.2 Intervention as product 50
4.1.4.3 Intervention as process 50
4.1.4.4 Validity, reliability and generalisability 51
4.2 Implementation
  4.2.1 Data collection
  4.2.2 Results
    4.2.2.1 Quantitative results
    4.2.2.2 Qualitative results

4.3 Evaluation
  4.3.1 Conclusions
  4.3.2 Recommendations for the English department and school
  4.3.3 Recommendations for further research and classroom application

CHAPTER 5: CRITICAL REFLECTION

5.1 Proceedings and reception of research

5.2 Effects and generative potential of research

SUMMARY (IN DUTCH) / SAMENVATTING

READING LIST

APPENDICES

1 Connectives chart
2 Profile of learner corpus
3 Pre- and post test (of design research)
4 Student questionnaire about effectiveness of intervention on connectivity and corpus linguistics
5 Intervention, sample materials
‘Digital research corpora for language, literature, philosophy, art and music have revolutionised the humanities. As digital storage and retrieval databases, these corpora enable researchers to discover patterns and themes that open up new avenues of scholarly investigation.’

- Inaugural address by Rens Bod, 6 October 2011, Universiteit van Amsterdam

1See http://www.english.uva.nl/news/professorappointments.cfm/3476083D-5913-4663-9E9F02DF06446A5A
This post-graduate paper reposes on my undergraduate research on compensating reading strategies and the significance of textual cohesion for intermediate (Dutch) students’ reading skills. Since then, my interests in applied linguistics have continued to develop in the fields of text linguistics, cohesion studies, corpus linguistics and their respective applications in English foreign language teaching, an example of which is found in this classroom-based research report.

In helping me complete this research project I should like to thank a number of people. First of all, my students at Sint-Maartenscollege in Voorburg who participated without bias in the various experimental activities I subjected them to and who provided the necessary linguistic data and critical feedback on the implementation of my research. Without them this study simply could not have been carried out. Secondly, my supervisor dr. Eline van Raaphorst whose sound and speedy comments on my findings and queries helped to minimise the luring delays so often facing ‘teacher researchers’. Next, a number of experts in the field of (applied) linguistics that I consulted and who kindly provided me with useful resources and ideas: Ramesh Krisnamurthy’s (lecturer in Corpus Linguistics, Aston University Birmingham) unrelenting enthusiasm for and immense experience with corpus linguistics have certainly paved further ways for the application of web-based corpora in the (Dutch) TESOL classroom; also, Mike Hannay (Free University Amsterdam), Ted Sanders (University of Utrecht) and Ildiko Berzlanovich (University of Groningen), who provided stimulating input at certain stages in my research.

It is my sincere hope that this project will find its way into the classroom and be a small stepping stone towards a more integrated approach to connectivity, corpora and data-driven (English) language teaching and learning in the Netherlands and, ultimately, contribute to more beneficial writing and language skills in general.

David Geneste, February 2012

---

2 See Geneste 1999.
1.1 Research context

An increasing number of secondary schools in the Netherlands are offering curricula in which traditional school subjects are taught partially or predominantly in English. Students in these so-called bilingual streams, and particularly those who are likely to attend international programmes and exams, are expected and trained to develop their receptive and productive English language skills to near-native level.

One of the most popular of such programmes is the international English Language and Literature diploma programme provided by the International Baccalaureate Organization (IBO). This a rigorous two-year pre-university course of studies in which Anglophone literatures, cultures and languages are studied.\(^3\) Assessment through written tasks, including comparatives, commentaries and essays, is particularly significant within this programme and successful candidates are expected to produce 'well-informed, effective and coherent discourse' appropriate to (pre-) academic writing.\(^4\) The IB programme is also offered to year five and six students at Sint-Maartenscollege (secondary school) in Voorburg, which is the professional context in which this research was carried out.

Another set of descriptors related to writing proficiency at advanced level is provided by the Common European Framework of Reference for Languages (CEFR).\(^5\) Discourse competence in both writing and speaking and the more specific aspects of 'coherence and cohesion' are essential requirements included in the CEFR. The global scale descriptors for proficient users (levels C1 and C2) require students to "produce clear, well-structured, detailed texts on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices (C1) and coherent and cohesive texts making full and appropriate use of a variety of organizational patterns and a wide range of cohesive devices (C2)".\(^6\)

In relation to foreign language teaching in the Netherlands the descriptors for writing proficiency are, predictably, global variations on the CEFR criteria. For writing at level B2(+), the highest described level in the Dutch interpretation and classroom implementation of the CEFR, the descriptors require that writers employ a limited number of cohesive devices so as

---

3 See http://www.ibo.org/ for general background information.
4 See International Baccalaureate Organization 2002: 44.
to produce clear and coherent texts with a logical organisation.\textsuperscript{7} Ironic in this respect is that a growing number of teachers of English in the Netherlands currently tend to deviate from the traditional set of guided writing principles and corresponding text genres (e.g. formal letters) at school examinations and instead stimulate and require students to produce freer written discourse (e.g. essays or reports).\textsuperscript{8} In the near future, this would certainly entail a reorientation on the teaching of writing skills and text genres.

From the criteria and descriptors for writing skills set by the IBO and the CEFR it becomes clear that the notions of coherence, cohesion and connectivity are essential aspects of written language production, especially at higher, pre-academic levels. In this study the concept 'cohesion' is loosely defined as the conceptual and linguistic links that hold a text together and give it meaning. Connectivity is one aspect of cohesion and observed as the sum of lexical devices that turn separate words, clauses, sentences and paragraphs into connected discourse, signalling the relationships between ideas and making obvious the thread of meaning the writer is trying to communicate on surface level. In this respect, connectives also constitute an important facet of 'cognitive academic language proficiency' (commonly abbreviated to CALP) i.e. the more formal and abstract language typical for (pre-)academic language output expected of advanced students in IBO and VWO courses.\textsuperscript{9}

Unfortunately yet not entirely surprisingly, connectivity is also one of the most demanding aspects of discourse competence for students to master and instructors to teach. For example, teachers of English commonly find that linguistic fluency comes quite easily to most Dutch students. At the same time, the same students regularly lack the capacity to choose the appropriate register i.e. distinguish between the use of formal and informal language and connectives, or between written and spoken discourse. Too often, then, this results in the use of informal connectives – [still], [anyhow] and [well] being the most obvious examples - in a formal writing context. This conclusion is supported by various researchers in (applied) linguistics. Crewe, for example, states that “The misuse of logical connectives is an almost universal feature of ESL students’ writing, though it may also occasionally happen with experienced writers.”\textsuperscript{10}

\textsuperscript{7} Fasoglio & Meijer 2007: 22, 47 and 52.
\textsuperscript{8} Kwakernaak 2009: 211.
\textsuperscript{9} See Dale, Es & Tanner 2011.
\textsuperscript{10} Crewe 1990: 317.
1.2 Research problem

The need for advanced skills in linguistic cohesion, of which connectivity is only one area, is a generally accepted assumption in the TESOL community these days. Mahlberg, a specialist in the field of text cohesion, states that ‘readers and writers should be aware of the links that hold chunks of text together and that contribute to the creation of a text as a unit of meaning. Cohesion can contribute to the readability of a text and have an impact on the comprehensibility and clarity of the argument.’ She continues: ‘An appropriate use of cohesive devices is essential for language learners to develop a native-like competence of text production and reception’.

In his extensive reference guide on the teaching of grammar Cowan observes that ‘accurate use of English discourse connectors is essential for TESOL students writing academic and technical English’. At the same time, TESOL learners express significant difficulties in mastering appropriate usage of connectivity. Overuse, underuse or misuse of connectives in English are common problems, leading to unconnected and obscured discourse. Some reasons for this include the (adolescent) learner’s inadequate linguistic and text-analytical competence, the influence or interference of mother tongue, cognitive restrictions and the nature of the instruction students have received and the input and materials they have worked with.

Crewe’s ‘universal feature’ substantiates the assumption that the majority of Dutch students of English find the effective use of writing of cohesive texts problematic, particularly at advanced level where a great degree of cohesion, connectivity and precision are desirable. And although methodological issues concerning the nature, problems and remedial teaching of cohesion and coherence have been given more attention in the past two decades, very little is known about the type and nature of mistakes in connectivity generated by Dutch students of English. This makes it extremely difficult for teachers of English in the Netherlands to make useful diagnoses and design remedial practice and to tackle this problem successfully. Insight into the scope and nature of errors in connectivity usage, therefore, would be beneficial to material developers, teachers and students alike.

---

12 Cowan 2008: 615.
15 Granger expresses the need for more large-scale empirically-based studies to obtain a more accurate view. See Granger and Tyson 1996: 18-19.
1.3 Objectives, scope and organisation of research

This study consists of two distinct yet closely related types of research. The first part, the analytical research, aims to show how connectives are used by Dutch adolescent learners of English at advanced level. More specifically, the analysis, based on a learner corpus of representative texts, will identify, categorize and describe connectives and inappropriate usage according to their syntactic, semantic and lexical properties. The idea is to diagnose which connectives the students find most problematic.

Analysing texts and categorising mistakes will give insight into the nature and extent of misuse. These findings, in turn, will serve as a linguistic and methodological basis from which to develop, try and assess relevant remedial teaching instruments and materials in the second stage of this project. The design research, which takes the recommendations in the analytical research as its starting point, is more practical in nature and has a classroom intervention at its basis. The aim of the design research is to see if and to what extent an innovatory approach to the teaching and acquisition of connectives can improve usage.

As we will see in chapter 2, connectivity is only one aspect of cohesion studies, which is a field too broad to be dealt in the scope of this research paper. As such, cohesion is restricted to that of connectivity because it is expected that the benefits of improved connectives usage by (Dutch) learners of English in generating cohesive texts are more substantial than that of other cohesive devices. At this stage it should also be said that contrastive language analysis, i.e. using features of the SL to account for inappropriate TL output, is not part of this study nor are cognitive-linguistic aspects of the target group, relevant and interesting features though these may be.

1.4 Abbreviations and terms used in study

The following abbreviations and terms (in alphabetical order) are used throughout this study.

BAWE: British Academic Written English (corpus)
CBLL: Content-Based Language Learning
CEFR: Common European Framework of Reference
CLIL: Content and Language Integrated Learning
COCA: Corpus of Contemporary American
DDLL: Data-Driven Language Learning
IBO: International Baccealaureate Organisation
**L1-learner**: first language learner

**L2-learner**: second or foreign language learner

**LC**: learner corpus

**NNS**: non-native speaker

**NS**: native speaker

**RC**: reference corpus

**SFP**: sentence-final position (of connective)

**SIP**: sentence-initial position (of connective)

**SL**: source language

**SMP**: sentence-medial position (of connective)

**TESOL**: Teaching English to Speakers of Other Languages

**TL**: target language
2.0 Introduction

This chapter presents the theoretical framework of this study. The focus is on the relationships between connectivity, corpus linguistics and TESOL. Specific theoretical and methodological issues relevant to the analytical and design research are dealt with in chapters 3 and 4 respectively.

2.1 Connectivity

2.1.1 Definitions and Research

Cohesion, coherence and connectivity are closely related terms and concepts which, by nature of this proximity, are often used interchangeably. Though a number of overlapping definitions exist, in this study cohesion is taken to mean the grammatical, and lexical relationships within a text, paragraph, sentence or clause that bind these units together and give them meaning. The seminal text on linguistic cohesion is Halliday & Hasan 1976 which identifies five categories of cohesive devices: reference, ellipsis, substitution, lexical cohesion and conjunction. A more recent article providing examples of these categories is found in Pander Maat & Sander 2006(a). Though all of these five categories are equally valuable in establishing and teaching cohesion in discourse, this study is solely concerned with what is commonly labelled as ‘conjunction’.

However, from a purely grammatical point the restrictive function of conjunction does not fully cover the type of cohesion as discussed in this context. Here, any type of lexical item connecting underlying relationships between stretches of discourse is interpreted as a ‘conjuncting device’ in the literal sense of the words. Terms used to denote such connecting devices, in both written and spoken discourse, include conjuncts, linking words, cohesive devices, linking adjuncts and discourse marker, the latter of which seems to have become more or less the most popular term in the field. These terms all have specific grammatical and to some extent overlapping properties but their interchangeable use causes considerable confusion in usage. Four sources that use different terminologies but all discuss the meaning and use of the same phenomena are Ball 1986, Chalker 1996 and the chapters on sentence connection in Quirk & Greenbaum 1973 and discourse markers in Swan 2005.

17 For a discussion of different terms denoting connectivity see Müller 2005: 3-4.
For the sake of transparency and convenience, this study attempts to categorize all of these lexico-grammatical terms and devices under the heading connectives and label the underlying concept as connectivity. This definition follows Pander Maat and Sanders’ in which all “connectives are one-word items or fixed word combinations that express the relation between clauses, sentences, or utterances in the discourse of a particular speaker” and, we should add, writer. From a receiver’s perspective, Celce-Murcia and Larsen-Freeman define connectives as “lexical expressions that may add little or no propositional content by themselves but [...] serve to specify the relationships among sentences in oral or written discourse, thereby leading the reader/listener to the feeling that the sentences ‘hang together’ or make sense.” In other words, when used appropriately with regard to meaning, syntax and style, connectives should function as lexical signposts, guiding the reader or listener through the discourse in the most logical and therefore most effective way.

2.1.2 Linguistic properties and learner problems
Nearly all connectives have individual and specific forms, meanings and functions, albeit that there are many that have common backgrounds and origins (e.g. [Further,] and [Furthermore,]). Connectives can be classified on the basis of general properties such as sentence positioning and style. These properties are discussed below and will prove useful for the analysis and understanding of individual connectives. Specific examples of inappropriate are discussed in see chapter 3.0.

2.1.2.1 Semantic properties and learner problems
Connectives are used to denote meaningful or semantically logical relationships between stretches of discourse i.e. words, phrases, clauses, sentences and paragraphs. There are many different ways of categorising connectives on the basis of such relationships but linguists that hold the ‘minimalist view’ generally agree on four major types - additive, adversative, causal and temporal - and these four categories are best represented by the four words: [and], [but], [so] and [then]. However, specialists that take the ‘maximalist view’ express the need to sub-categorize and further define these ‘conceptual primitives’ to thus express more subtle degrees in connectives usage.

18 Pander Maat & Sander 2006 (b).
20 This categorisation is drawn from Hasan & Halliday 1976.
21 See Pander Maat & Sanders 2006: 34 and Evers-Vermeul 2005: pp. 13-20 for a fuller discussion of the nature of these conceptual primitives.
Two writing guides frequently used in the Netherlands perfectly illustrate the extremes of the ‘minimalist’ and ‘maximalist’ views on the range of relationships to be offered. Hannay & Mackenzie 1996 limit their categories to a mere four and gear these to specific composition writing needs. Laaken, Lankamp & Sharwood 2001, on the other hand, supply a classification of no fewer than thirteen categories. And although such elaborations may be of interest and value to relatively experienced writers they often cause confusion as logical categories and corresponding connectives are drawn from different sources.\textsuperscript{22} To make matters even more complicated, some lexically identical connectives (i.e. [then]) express different logical relationships (in the case of [then] these are causal and temporal). Therefore, this study concentrates on the aforementioned four conceptual primitives. See Appendix 1 for an overview of categories and corresponding connectives used in this study.

Recognising and using semantically correct connectives is a demanding task to L2 learners and often leads to inappropriate use of connectivity. The most common of interrelated problems are: a lack of understanding of the underlying semantic concept (e.g. causality); a lack of knowledge of connective resulting in underuse; and inappropriate use of connective resulting in misuse.

\textbf{2.1.2.2 Syntactic properties and learner problems}

In addition to conveying specific semantic properties, connectives also express specific syntactic properties i.e. specific positional features of the connective in the discourse unit, most often in the forms of a sentence or clause. Generally, there are three possibilities: (1) sentence-initial position (SIP); (2) sentence-medial position (SMP); (3) sentence-final position (SFP).\textsuperscript{23} The majority of connectives in English take sentence-initial position but some can only be used in sentence-medial or final position depending on the restrictions posed by the specific syntactical features of the discourse unit.

An additional characteristic closely related to positional features of connectives is the use or lack of use of commas indicating the position and highlighting the specific function of the connective in question. Unfortunately, the inconsistent use and positioning of commas in L1 discourse provides relatively little reliable input for L2-learners and continues to be a persistent source of confusion to L2 learners.

Positional problems with connectives are caused by L1 interference, the mixing up of connectives that express different functions but have fixed positions, and, most importantly,

\textsuperscript{22} Examples of subcategories of logical concepts can be found in Quirk & Greenbaum 1973 and Ball 1986.\textsuperscript{23} For a discussion of these syntactic primitives, see Evers-Vermeul 2005: pp. 20-29.
the overuse of sentence-initial position. Generally, L2 learners are inexperienced in manipulating connectives with sentence structures. In addition, L2 learners tend to underuse commas, whose primary function it is to stress the position and function of connectives.

2.1.2.3 Stylistic properties and learner problems

Style, or register, in the context of this study is meant to be understood as the usage of connectives appropriate to formal writing (i.e. analytical essays), as opposed to formal or spoken discourse. This is because a number of connectives are style-sensitive and can only be used in formal or informal contexts.

Problems arise when L2 learners are not familiar with the specific register of the text genre or usage of connectives. This is particularly the case with secondary-school students who have relatively little experience in writing formal or academic English and are prone to confuse formal connectives with colloquially-marked such as [anyway], [so], [actually].

2.1.3 Connectivity teaching: current practices and restrictions

As stated above, connectivity has for some time been considered an important item in language learning and teaching. As such, it is often presented in textbooks on grammar or writing skills, particularly those used at (pre-) academic level. In the latter category the various aspects of connectivity are presented as one interrelated field of formal language analysis and acquisition whereas in grammar guides connectivity is dissected and distributed over various fields of traditional grammatical analysis. The lexical-functional approach addresses connectives as lexical items i.e. as fixed idioms rather than as grammatical forms which can be variable and manipulated. The standard format of presentation in a number of such books is through lists of connectives, categorized according to their semantic relationships. In the corresponding set of exercises learners are to identify, choose, correct or fill in the appropriate connective in a given and often limited set of sentences or contexts.

24 See Granger & Tyson 1996: 25.
26 Chalker 1996 is the most ambitious and complete textbook in presenting ‘linking words’ as one coherent field of English foreign language study. Murphy 2004, one of the most widely used grammar book in the field, covers connectivity in various disparate units on e.g. conjunctions and prepositions.
27 See Mahlberg 2006: 228.
28 The classification in Quirk & Greenbaum 1973 is the standard reference for categories of connectives. The connectives chart in Appendix 1 of this study is based on Quirk & Greenbaum 1973.
These existing and widely-used approaches to connectivity teaching have obviously served useful purposes. However, with a view to the design research of his study a number of alternative methods should be provided here. Firstly, connectives should always be introduced as higher-level discourse units aiming at producing connected and logically organised texts and not merely as ‘stylistics enhancers’, as they sometimes are.\(^{29}\) As such, they should be integrated into advanced TESOL course components.

Secondly, lists and categories of so-called interchangeable connectives presented can lead to inappropriate usage.\(^{30}\) A textbook may present words such as [in addition], [furthermore] and [moreover] as simple additive connectives that have the general meaning of [too] and [also]. The result is a lack of appropriate and ample exemplification, which is essential to illustrate and understand the complexity and subtleties of connectivity. If the differences between them and the contextual conditions that favour the choice of one or the other are not provided, there is nothing to prevent the student from believing that these connectives are interchangeable.\(^{31}\) Instead, connectives should always be presented in authentic reading and writing contexts, preferably at text rather than at paragraph or sentence level, rather than in isolated fragments and exercises. Again, it is better to have one authentic context with fewer connectives that constructed texts with more. In this respect, the aforementioned writing textbooks are much favoured to the grammar books as a means of presenting and practising connectivity. Moreover, effective use of corpora can help students a great deal, as we will see later on.

**2.1.4 New starting points and directions in connectivity teaching**

The aim of this study is to improve students’ use of connectives in writing contexts. The focus, therefore, is on the productive (writing) skills but we should not forget that the recognition and understanding of connectives is equally important. It is essential, therefore, to place emphasis on how connectives are used semantically, syntactically and stylistically in authentic texts, as was stated before. This can be done by examining their use in authentic texts and teaching students where connectives should or should not have been used. It is imperative that students develop profound analytical reading skills on relevant linguistic levels ranging from word, phrase, clause, sentence and paragraph to essay level. This can be achieved through extensive reading and close analysis of various kinds of text genres.

\(^{29}\) Granger 1996: 25.
\(^{30}\) Granger 1996: 25.
\(^{31}\) See Cowan 2008: 631.
Another relevant point relates to connectivity and texts approached bottom-up or top-down. It is obvious that both approaches have pros and cons but, ideally, students are trained to always place and interpret isolated language chunks in larger and more authentic contexts. Therefore, it is important to illustrate the function of connectives in relation to larger linguistics units of meaning, as discussed before. On the other hand, the often complicated semantic, syntactic and stylistic nature of many connectives requires a closer, bottom-up analysis of these properties. A relevant and balanced combination seems to fit our purpose best.

Next, students should be familiarised with the nature and purpose of connectivity in relation to what they are expected to do with a view to their educational careers. In addition, some but not too much theory on the nature of connectivity should therefore be provided. For example, students should be able to understand the three basic linguistic properties of connectives – semantic, syntactic and stylistic – if they are to recognize and (self-)remedy inappropriate use. Theoretical discussions about what is considered a connective and what is not, on the other hand, are best avoided to avoid confusion.

Another relevant issue is concerned with the number of connectives or degree of connectivity students are to use appropriately. In other words, are learners expected to master a limitless choice of (lists of) connectives or should the focus be on a set and limited number of them? And if the latter approach is preferred, what criteria are to be used? Textbooks often contain too few constructed sentences to illustrate the complexity of connectivity. Of course, time is limited and it is nigh impossible to find useful and authentic texts that contain enough connectives to serve teaching purposes. In addition, a detailed textual analysis is time-consuming: the analysis of full texts required more attention that the discussion of just a few sentences. Generally, an in-depth exploration of a limited number of contextualised connectives is preferred.

With respect to types of connectives, the text type in question can provide a solution as different types of text genres require different types of connectives thus conveying a greater or lesser number of a certain category of connectives. In the corpus analysis of this study, for example, literary essays were used as input. It was expected that students were likely to use illustrative connectives (e.g. [for example/instance] more frequently than other connectives as they were explicitly required to integrate examples of textual evidence into their arguments. Alternatively, the text genre used in the intervention of this study is the

---

comparative commentary so it is logical that more attention is paid to adversative connectives as learners are expected to contrast and compare ideas and arguments. A focus on connectives posing specific problems to Dutch learners (or those of other first-language backgrounds) may be another way to select a specific and limited group of connectives. A survey of these may be found in contrastive textbooks on writing such as Hannay & MacKenzie and Lemmens & Parr.\(^{33}\) This method, however, implies a contrastive approach to TESOL.

On a final and perhaps superfluous note it should be said that connectivity, as part of the mechanics of cohesion, remains only one way of achieving cohesive texts. To this may be added that cohesion in written and spoken discourse have differences and similarities, the mutual interaction of which may be beneficial to the improving of connectivity skills in general. Ideally, then, other aspects of cohesion should also be touched upon with in a TESOL course on writing skills.

2.2 Corpus linguistics

2.2.1 Definitions and research

Corpus linguistics is a relatively new branch within linguistics and is concerned with the purposeful compilation of extensive digital text files – named corpora – which can be analysed using specific software and search tools.\(^{34}\) The Internet, for example, could be considered a global corpus with Google as a search tool. Unfortunately, the tools available do (as yet) no allow linguists to look for specific linguistic utterances that meet accepted and verifiable standards. Professional corpora, however, consist of authentic and verified language usage based on huge quantities of written and/or spoken language output using a set of fixed criteria such as text genre and size and learner background. Linguists interested in connectivity, for instance, can investigate how the connective [however] is used in ‘real’ language by ‘real’ language studying a representative collection of linguistic data, either compiled as written texts or as a transcription of recorded speech.

An example of such a very large corpus is The Corpus of Contemporary American English (COCA) which is made available through Brigham University in the United States. \(^{35}\)

\(^{34}\)For a readable introduction to corpus linguistics see Teubert & Čermáková 2007 and McEnery, Xiao & Tono 2006.
COCA currently consists of 425 million words and is freely accessible for researchers, teachers and students. Additional advantage is that it has integrated search tools allowing for relatively easy analyses, which is why it is used in the intervention of this research.

The main purpose of a corpus search is to verify a hypothesis about language, for example by determining how connectives are used ‘naturally’ by a representative group of native-language or foreign-language users. To this end, researchers can use existing on-line corpora or construct their own corpora using the input from a select group of learners. A recent relevant large-scale contrastive corpus analysis of this kind was carried out by Tavecchio who has investigated to what extent syntax and text genre determine differences between Dutch and English sentence-patterns.\(^{36}\) From a pedagogic point of view, the corpus-based research by Springer is particularly interesting as it looks at various features of the discourse competence of advanced Dutch writers of English, including connectivity, from a contrastive point of view.\(^{37}\) For the analysis in this paper a learner corpus was compiled and analysed using on-line text-processing software such as TextStat. This learner corpus was then analysed to see how often (quantitative) and in what way (qualitative) connectives were used in by the learners. More details are to be found in chapter 3.

Corpora were initially used by linguists, translators and lexicologists. For example, the context sentences in the renowned English Cobuild dictionaries were extracted from the Bank of English, a large scientific corpus designed at the University of Birmingham in the 1980s. Gradually, however, corpus linguistics is finding its way into the TESOL classroom.

### 2.2.2 Corpus linguistics and TESOL

Corpus linguistics is rapidly expanding field within linguistics and it has developed its own theories and tools in recent years. Internet access allows researchers (and students) to use, research and develop corpora in increasingly accessible ways.\(^{38}\) Its pedagogic applications, however, are only slowly being applied in the diagnosis of language use (error analysis) and the design of remedial materials. Two reasons for this may be that both ELT teachers and students have ‘a preference [...] for straightforward right-or-wrong answers, rather than the sometimes fuzzy evidence that corpora provide’ and ‘working with corpora may not suit some learning [or teaching] styles’.\(^{39}\) Consequently, corpus-driven activities, tools and materials

---

\(^{36}\) See Tavecchio 2010.

\(^{37}\) See Springer 2012.

\(^{38}\) See Anderson & Corbett 2009 for ideas and sources.

\(^{39}\) See Marks 2010 for other reasons for the slow acceptance of corpus linguistics in ELT.
such as the ones used in this study are to some extent experimental and their use depends on the willingness or expertise of the teacher.

Despite the doubts the ELT community may have about the classroom applications of corpus linguistics there are a growing number of publications advocating the methodological advantages of corpus linguistics and practical classroom ideas. Corpora can also be used to convey particular linguistic data visually making systems and frequencies in language and texts more accessible to (adolescent) language learners. An example of such a creative manipulation of quantitative data can be found on the cover of this research report.

Teachers can benefit from corpora from two specific methodological perspectives. Firstly, they can compile learner corpora using their own students' language input that is subsequently processed and analysed using web-based software and tools. In doing so, they can carry out action research and focus on areas or learner errors specific to their own situation. Secondly, existing on-line corpora can provide valuable language input for students, for reference of revision purposes. Teachers can use corpora as language input for the development of material and activities. Both perspectives and uses of corpora are explored and assessed in this study. In addition, the TESOL community seems to agree that corpora offer two major linguistic advantages: language items used in context and in an authentic i.e. natural way, and enabling learner involvement. These principles have for a long time been regarded important factors in foreign language teaching.

As to the issue of learner involvement, a specific view on language teaching and the use of corpora is particularly interesting. The notion of ‘data-driven (language) learning’ (DDL) was developed by Tim Johns in the 1990s. The basic idea is that learners are encouraged to work out their own rules about the meaning and patterns of usage of particular linguistic units. Learners are trained to use a so-called concordancer tool which helps them to locate examples in a corpus of authentic texts. Corpora, of course, are extremely suitable for this purpose as they never provide ready-made answers but, instead, give the inquisitive and independent language learner the opportunity to locate and deduce general linguistic patterns based on authentic language usage. In this way, the corpus can become a useful type of tool for further language development and autonomous learning.

---

40 See Carmen, Lluisa & Belles-Fortuna 2010 for research exploring different ways to apply corpus-based approaches to ELT and O’Keeffe, McCarthy & Carter 2007 for classroom applications. See Bennet 2010 for classroom ideas.
41 The word-cloud picture on the cover was created using www.wordle.net, an online programme which transforms a corpus i.e. this research paper, into a word cloud. The word cloud visualises the frequency of words through the prominence of the word in the cloud. See Scott & Tribble 2006.
42 See Boulton 2010 and Shaw 2010.
2.3 A corpus-based approach to connectivity acquisition and teaching

In concluding this chapter we have to address the issue of how connectivity and corpus linguistics can be combined in the TESOL classroom. The area of linguistics known as learner corpus research suits the needs for research into the analysis of connectivity usage and the design of remedial materials well. Computerised data can be analysed with linguistic software tools which search, count, display and even analyse the data.

As briefly stated before, the connectives used in this study are drawn from two types of (written) computerised corpora, a learner’s corpus (LC) and a reference corpus (RC). Learner corpora contain large quantities of discourse to be investigated produced by a select group of non-native speakers (NNS). Reference corpora, on the other hand, contain discourse generated by native speakers (NS) against which the learner corpora and corresponding language queries are assessed. Corpora can be compiled using traditional personalised classroom data or anonymous online data, each with its own specific features and merits.

In recent years a number of online corpora have been made available which can be consulted and used for the study of language at different linguistic levels. In addition to using existing online corpora researchers and teachers can also compile their own corpora using (free) internet sources such as online essays, dissertations and research reports permitting appropriate criteria or learner and corpus profiles are used. Text-processing software allows for fast and systemic generation of frequencies or concordances of e.g. connectives, which can then be analysed quantitatively (in isolation, using computer software) and qualitatively (in context, analysed manually). In this study a number of both LC and RC corpora have been used.

The intervention in the design research of this study distinguishes between receptive and productive phases. The focus of the receptive phase is clearly on students finding out about the appropriate or inappropriate use of connectives in reference or learner corpora. In this phase they are also first introduced to connectivity, corpus linguistics and its relation and relevance to language learning. They will learn to use digital corpora in the same way that other ELT tools help them (e.g. a dictionary or grammar reference). In the productive phases, then, learners are to use connectives themselves, first in relative isolation (i.e. in individual or sets of sentences) and then in a full context (i.e. a composition). In the concluding phase learners will use corpora again, this time to revise their own use of specific connectives by

43 For an overview of learner corpus research see Granger 2002. For an application in grammar teaching see Meunier 2002.
44 See Lew 2009 for the pros and cons of traditional and online corpora.
45 See Anderson & Corbett 2009.
checking these against the use of native language users. The various steps of this procedure are explained in detail in the description of the intervention.

Finally, teachers can also use corpora to develop needs or topic-specific classroom materials and activities. With reference to connectivity, for example, it is particularly easy to locate a great number of verified example sentences denoting subtle differences in usage of certain connectives. The exercises used in the pre and post tests of the intervention of this design research have been developed this way.
3.0 Introduction

Central to this chapter is the analysis of the learner corpus and connectivity usage. In addition to a discussion of the research proceedings and findings, specific background theory is presented below.

3.0.1 Research into connectivity usage and analysis

Traditionally a stronghold within Translation Studies, the study of the theory and application of connectivity in English discourse has grown significantly in the past twenty years, particularly in areas where a reorientation on the learning and teaching of English as a foreign language has become a pressing demand, in Eastern Europe and South-east Asia in particular. The objectives have been and continue to be to find out about language-specific connectivity usage and the challenges this generates, and the design of courses and materials for effective remedial practice. Remarkably, in-depth research into the usage of English connectives by Dutch learners, from contrastive or monolingual perspectives, is scarce. Sanders 1994 states that research into connectivity in Dutch as L1 is highly desirable. His wish has been fulfilled to some extent by dissertation research into the development of Dutch connectives by Evers-Vermeul 2005 and Springer 2012.

It should be repeated at this stage that a contrastive analysis of connectivity in Dutch and English, which exceeds the scope of this paper, is expected to yield interesting results with respect to remedial teaching and practice. If conclusions are to be drawn about the use and nature of connectivity in both languages, large-scale quantitative analyses of written discourse in English and Dutch have to be conducted. One pressing question in this context, for example, is to determine the frequency of usage of connectives in Dutch and English and preferred choices. Tavecchio's corpus-based contrastive analysis of English and Dutch sentencing patterns provides useful new insights with respect to this.

The linguistic data and outcome of the ample studies carried out in China and Eastern-Europe are mostly language-specific but some general inferences and methodologies employed in these studies are applicable to the Dutch situation as well. One aspect that is particularly useful to our purpose is the introduction of (digital) language corpora enabling

---

46 For the relationship between cohesion and Translation Studies, see Baker 1992.
47 See the introduction of Bikelenè 2008 for recent studies.
48 See Tavecchio 2010.
linguists to analyse large-scale surveys of authentic language input. In this way very broad yet far-reaching conclusions may be drawn and used for further research. This study also generates a corpus of digitalised texts and a processing programme based on a standard methodology designed to categorize and analyse linguistic data; more information is found in units four and five.49

Nearly all of the studies into connectivity state that three aspects are important when analysing connectives in written texts. These are: 1) the identification of the connective through form and function; 2) the linguistic features of the connective from a syntactic, semantic and stylistic perspective; and 3) the specific obstructions the connective causes through misuse, overuse or underuse. As misuse is the most interesting type of obstruction, or inappropriateness, in connectivity, this category will be further analysed on the basis of semantic, syntactic and stylistic features.

3.0.2 The identification and labelling of connectives

Each connective consists of a form, i.e. the surface discourse, and a function, i.e. the relational or logical concept it represents. It is unfortunate for many learners of English (and many other languages) that the relationship between form and function is completely arbitrary making memorisation and usage complicated.

Numerous exhaustive surveys of English connectives exist, each of which is equally valuable depending on the purpose it is used for. The thirty connectives in this study have been drawn from Hannay and Mackenzie’s Effective Writing in English.50 Though primarily meant for higher-education students with advanced proficiency in English, this writer’s resource guide contains extensive notes on the use (and misuse) of connectives by Dutch learners and can be used for our target group as well.51 In addition, a number of connectives not in Hannay’s and Mackenzie’s list but based on experience in analysing student essays and connectivity have been added. These combined lists of connectives should serve as a reliable reference tool for analysis and thus give a clear and relatively comprehensive idea of the use of connectives by our target group.

49 For the advantages and disadvantages of using traditional and web-based corpora see Lew 2009.
3.0.3 Inappropriateness in connectivity

There are three categories of inappropriate use of connectives: underuse, overuse and misuse. This study concentrates on the misuses of connectives as these surface data are most suited for computerised quantitative analysis. The categories are illustrated below using a standardised extract that has been manipulated for each of the three categories.\(^{52}\)

The original connectives used are [however], [even so] and [until] and can be found in extract 0 below, which is an example of appropriate use of connectives. When used appropriately connectives link one concept or sequence to another, providing a logical sequence of ideas and easy flow of understanding and reading. Appropriate use is typical for experienced language users with a firm and natural grasp of connectives.

Original extract 0: **Appropriate use of connectives**

*The Princess avoids publicity*

*The Princess was taken to another access point, which sent photographers crashing through the rain forest like a panicking herd of elephants. Even so, the photographers kept their word and not a shutter was clicked until the Princess took up the pre-arranged position. The mood, however, was distinctly frosty.*


3.0.3.1 Underuse of connectives

In underuse a connective is not used where it should have been used to clarify the relationships between two units of discourse. The result is an illogical, unnatural and therefore difficult to read sequencing of discourse units. Underuse is typical of learners with little experience in and knowledge of cohesion and connectives. Underuse is difficult to diagnose in a computerised quantitative analysis as surface elements are not present and therefore difficult to identify. As such, the analysis of underuse is not included in this study. Extract 1 is representative of discourse in which appropriate connectives are not used.

---

\(^{52}\) Extract taken from Ball 1982: 50-51.
Manipulated extract 1: Underuse of connectives

The Princess avoids publicity
The Princess was taken to another access point, which sent photographers crashing through the rain forest like a panicking herd of elephants. The photographers kept their word and not a shutter was clicked. The Princess took up the pre-arranged position. The mood was distinctly frosty.


3.0.3.2 Overuse of connectives
In overuse a connective is used where cohesion has already been achieved satisfactorily. The result is an overdose of connectives, with overlapping or, in the worst case, conflicting properties. Overuse is typical of learners with a passive knowledge of connectivity but with little practice in applying them appropriately. Sometimes, (inexperienced) learners want to impress readers by using many 'difficult' connectives without realising their superfluity. A diagnosis of overuse requires a qualitative analysis of connectives. Overuse can also be classified under misuse.

Manipulated extract 2: Overuse of connectives

The Princess avoids publicity
Next, the Princess was taken to another access point, which sent photographers crashing through the rain forest like a panicking herd of elephants. Even so, the photographers kept their word and consequently not a shutter was clicked until the Princess took up the pre-arranged position. Then the mood, however, was distinctly frosty.

3.0.3.3 Misuse of connectives

Misuse of connectives is based on the inappropriate use on specific linguistic or conceptual grounds. This result is an illogical connecting of discourse units obstructing an easy flow of understanding and reading. As discussed before, connectives have semantic, syntactic and stylistic properties that may cause inappropriate usage. These are illustrated and explained here.

Manipulated extract 3: Misuse of connectives based on semantic, syntactic and stylistic properties

*The Princess avoids publicity*

*The Princess was after that* taken to another access point, which sent photographers crashing through the rain forest like a panicking herd of elephants. The photographers kept their word *moreover* and not a shutter was clicked until the Princess took up the pre-arranged position. *Anyway* the mood was distinctly frosty.


**Syntactic misuse:** The temporal connective [after that] in this sentence can only take sentence-initial position, followed by a comma. If the sentence had not included the subclause the connective could also have been placed in sentence-final position.

**Semantic misuse:** The use of the additive connective [moreover] is inappropriate because the relationship between the two sentences is not based on addition but on concession, which indicates that the following sentence is contrary to expectation (as the photographers were expected to start photographing without waiting). An alternative connective in this context would be [even so].

**Stylistic misuse:** Though the use of the concessive connective [anyway] may be semantically correct in this extract it is stylistically inappropriate because of its colloquial nature, which unsuitable for this text genre. A correct connective here would be [however]. In addition, the connectives should have been followed by a comma to accentuate its place and function.
Obviously, semantic, syntactic and stylistic misuses of connectives provide the most useful information and therefore constitute the bulk of this analysis.

3.1 Procedure
This chapter describes the objectives, research questions, plan and instruments used to conduct the analysis.

3.1.1 Objectives
The main objective of the analysis is to find out how connectives are used by a specific group of Dutch learners of English at advanced level. Through a careful planning and conditioning of the research setting, the results are intended to be representative for a much larger group of students in the Netherlands. The outcome and conclusions are to be used as the starting points for the design research and corresponding intervention, and the development of effective teaching tools and remedial practice materials.

3.1.2 Research questions and conceptual model
The objective of the analysis is to find out how English connectives are used by Dutch learners. The hypothesis is that learners show a significant degree of inappropriate use.

The primary research question can be formulated as:
What semantic, syntactic and stylistic inappropriateness in the use of connectives by adolescent Dutch learners of English at advanced level is present in their expository essays?

Secondary research question 1 (quantitative):
Which ten connectives are most frequently used by adolescent Dutch learners of English at advanced level?

Secondary research question 2 (qualitative):
Which five of the ten most frequently used connectives show misuse in comparison with native speakers’ use of connectives, and what are the semantic, syntactic and/or stylistic properties of the misuse of these five connectives?
Secondary research question 3:
*Which connectives and types of misuse are most obstructive in establishing appropriate connectivity and which require most attention in remedial teaching?*

Variables and conceptual model
The variable in question is the use of connectives and, more specifically, the use of semantic, syntactic and stylistic properties of connectives. It is expected that the use of connectives by the target group will result in misuse. The aim of this study is to diagnose the nature of this misuse using set criteria and instruments. From this assumption the conceptual model below follows.

<table>
<thead>
<tr>
<th>USE OF CONNECTIVES</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
</tr>
<tr>
<td>MISUSE OF CONNECTIVES</td>
</tr>
</tbody>
</table>

Generalizability and reliability
One of the aims of this study is to provide a model that can be used to analyse, diagnose and remedy inappropriate use of connectives in the English foreign language classroom in the Netherlands. Generalizability, therefore, is highly desirable but not always easily achieved. The background of learners and teachers, for example, may have far-reaching implications for the outcome. In order to minimize unreliability, potential weaknesses are discussed under each of the research instruments in chapter 5.

3.1.3 Research plan, conditions and characteristics
The steps below describe the analysis, conditions and characteristics of the research. For definitions of key concepts see the chapter on research instruments.

The language input is drawn from 26 Dutch secondary-school students, 12 male and 13 female, aged 15 or 16 and currently in their fourth year of 'VWO'.53 They are generally motivated for English and school in general and have followed a three-year course in bilingual education attending six subjects taught primarily in English. The students’ general writing proficiency at the CEFR would at this stage be assessed as B2/C1. Considering the age group it is not unlikely to assume that the majority of the students find linguistic as well as

53 'VWO' denotes the most theoretical of levels in Dutch secondary education preparing students for higher education.
cognitive aspects of connectivity and cohesion in general in both TL2 and TL1 demanding. Cognitive limitations specific for this age group and inevitable inappropriateness in connectivity should certainly be taken into account in any analysis of connectivity usage and design of materials.

The students' instructor is a near-native speaker of English who emphasises essay-writing skills and the importance of cohesion, more so than is generally done at this level. The students have been trained in a non-contrastive/immersive way. This is expected to influence the outcome of the analysis in a positive way in comparison to students who have attended a contrastive/non-immersive course in English.

**Step 1:** Thirty suitable connectives are used for the quantitative analysis which will yield raw frequency lists based on the LC and the RC. Students were required to write essays on a novel of their choice using an extended essay question. They had 90 minutes to finish the essay, using the word-processing programme *Word for Windows*. A spelling checker was allowed minimising spelling mistakes in the use of connectives. The text genre investigated in this study, the expository (analytical) five-paragraph (literary) essay, is expected to be cohesive and contain ample connectives. More so, students are required to include quotes and references from the source texts to substantiate their ideas. Consequently, a higher frequency of the example connectives [for instance] and [for example] is expected, the frequency of which should be taken into account in the results.

**Step 2:** From the thirty connectives the ten most frequently used connectives in the LC will be selected and analysed qualitatively.

**Step 3:** On the basis of the ten most often connectives the frequency and percentage of misuse is established. The five connectives with the highest percentage of misuse are selected.

**Step 4:** The five connectives misused most often are analysed with respect to their semantic, syntactic and stylistic properties using the concordances and context sentences generated by the text-processing software, and theory about connectives and connectivity. The following example will demonstrate this procedure:
The Princess was taken to another access point, which sent photographers crashing through the rain forest like a panicking herd of elephants. **[Though]** the photographers kept their word and not a shutter was clicked until the Princess took up the pre-arranged position. [...] ‘The Sunday Telegraph’, November 1982.

Inappropriate connective use in this passage is identified by [Though,] on the basis of syntactic inappropriateness. [Though] in this sentence takes sentence medial (or final) but not sentence initial position because it is conceded that the photographers in the previous sentences kept their word where other behaviour was or could be expected. [Though] in sentence initial position concedes a point of prediction in the subsequent sentence.\(^{54}\)

On the basis of the analysis of both the essays and the contextualised essay, a list of the five most frequently misused connectives will be compiled, on the basis of which the research questions will be answered and conclusions and recommendations are provided. These, in turn, will serve as a starting point for remedial materials and teaching.

### 3.1.4 Research instruments

**Research instrument 1: Reference connectives used for quantitative analysis**

Thirty reference connectives have been selected for the quantitative analysis. This selection is made bearing a number of criteria in mind. Firstly, it is expected that the selection is representative for the language level of the target group. This means that the relatively basic connectives [and], [or] and [but] have been left out. Secondly, some connectives are predicted to yield interesting results because they come from a list taken from Hannay & Mackenzie 1996 containing connectives that cause contrastive (English-Dutch) language problems. In addition, the connectives should be more or less univocal in meaning thus avoiding ambiguities and time-consuming searches in the raw frequency lists. For example, the word [as] has not been chosen because it is a connective with four different and context-dependent functions (i.e. adversative: [Much as ...]; temporal [As I was walking ...]; referential [As for ...]; and causal: [As it rained I ...]). It has to be emphasized here that the number of

---

\(^{54}\) See Ball 1986: 113.
reference connectives opted for is by nature selective and to some extent arbitrary. This is because the scope of this project does not allow for a more comprehensive analysis of connectives.

Research instrument 2: Reference connectives to be analysed for qualitative analysis
From the list of thirty reference connectives, the five connectives that convey the highest percentage of misuse are selected for further qualitative analysis. This shortlist will serve as a starting point for further research and the design of materials based on the analysis of students' corpus of analytical essays and their frequency of usage.

Research instrument 3: Learner corpus
The learner corpus (LC) consists of 25 analytical essays amounting to 14,866 words, generated digitally by 25 students from the same cohort. The essays have not been adapted or annotated. See appendix 2 for the profile of the learner corpus.

Research instrument 4: Reference corpora
In this study two reference corpora are used. The first corpus, named 123helpme, consists of 23 literary essays comparable in size and template to the ones produced by the L2 learners. The reference essays have been drawn from an online and public databank containing analytical (literary) essays, generated by native speakers with backgrounds similar to those of the learner corpus i.e. American English high-school students.55 The essays selected show consistency concerning language fluency, style and structure. However, as little background information is available about the specific background of the contributors and requirements for acceptance caution is in place as to the reliability of the references corpus.

The second corpus used is the open-access British Academic Written English (BAWE) containing 2761 pieces of proficient assessed student writing, ranging in length from about 500 words to about 5000 words, with a total of 6,506,995 words.56 The combination of the two corpora 123helpme and BAWE against which the LC is assessed should provide ample and reliable results.

56 The BAWE corpus is a collection of student assignments of reliable standards and drawn from undergraduate to postgraduate levels and from various disciplines. Further details of its organisation can be found at http://www.coventry.ac.uk/bawe Go to http://ca.sketchengine.co.uk/open/corpus/bawe2/ske/first_form for the tools required to use the corpus.
Research instrument 5: Text processing software for learner corpus

The open-source software used for this purpose is TextSTAT, which has been developed by the department of Dutch Studies at the Freie Universität in Berlin.\textsuperscript{57} TextSTAT is a programme for the computerised analysis of texts reading plain text files (in different encodings) and HTML files (directly from the internet) and producing word frequency lists and concordances from these files. TextStat makes text statistics and counts characters, words and sentences to find words repetitions. Other (open) text-processing programmes are available but TextSTAT is particularly suitable for our purpose in that it locates concordances of connectives in context in a simple and straightforward way.

3.2 Implementation

3.2.1 Data collection

In this chapter the implementation of the research is described. To collect the data, i.e. to locate the connectives most frequently used by learners, the learner corpus has to be assessed against the two reference corpora through extraction of the thirty selected connectives. The concordances in the learner corpus and reference corpus 1 are located using TextStat; the concordances in reference corpus 2 are found using Sketch Engine, which is the set text-processing programme used for the analysis of BAWE.

A note on the size of corpora is in place here. Ideally, learner and reference corpora are of equal size. In our case the size of the LC should amount to approximately 15,000 words. Reference corpus 1, consisting of web texts, is compiled in such a way that it contains approximately the same number of words as the learner corpus. Reference corpus 2, however, is compiled by a research institute and contains 6,506,995 words. It is, as such, significantly larger than either of two other corpora and cannot be adapted. Instead, the found concordances need to be adjusted to the maximum number of words in the learner corpus to make the corpus applicable to our purposes. In practice, this means that each finding from RC2 needs to be divided by 434.\textsuperscript{58}

As stated before, one of the reasons for compiling this specific list of connectives was to avoid ambiguity in usage and analysis. However, due to the nature of the text-analysis programmes all connectives have to be checked manually to see if they fit the criteria set in this context. For example, concording the item [again] also yields broader and in this context

\textsuperscript{57} See \url{http://neon.niederlandistik.fu-berlin.de/textstat/}
\textsuperscript{58} RCII 6,506,995 words divided by LC 15,000 words gives 434.
irrelevant items such as the preposition [against]. And, depending on the specificity of the research, it may also be wise to distinguish between the various meanings or uses of potentially ambiguous connectives. The connective [however], for instance, can denote two (similar) relationships, the first one being concession and the second one a concessive complement to an adjective, as in ‘However old you are, you can always start a new hobby’. In this limited research context such relatively subtle differences are not accounted for.

The procedures above are used for the quantitative analyses aimed at finding out how frequently a connective is used in relation to the referenced corpora. The qualitative analysis, that is the extracting of connectives used inappropriately, has to be done manually as has the labelling of the type of inappropriateness.

### 3.2.1 Results

In this chapter the results of the analysis are presented in relation to each of the research questions formulated in chapter 3.1. The table below, containing an alphabetical list of the thirty connectives and their raw frequencies in each of the corpora, will serve as the basis for the quantitative results.

<table>
<thead>
<tr>
<th>Connective</th>
<th>LC</th>
<th>RC1</th>
<th>RC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 accordingly</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2 again</td>
<td>11</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>3 also</td>
<td>28</td>
<td>31</td>
<td>39</td>
</tr>
<tr>
<td>4 although</td>
<td>1</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>5 anyway</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 because</td>
<td>41</td>
<td>38</td>
<td>16</td>
</tr>
<tr>
<td>7 besides</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8 concerning</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9 conclusion</td>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>10 consequently</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>11 especially</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>12 example</td>
<td>11</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>13 finally</td>
<td>8</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>14 firstly</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>15 furthermore</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 1 shows that the corpora have considerable consistency in the frequency of use of connectives yet also containing some interesting differences. In relation to RC2, the connectives [again] and [because] are overused in both LC and RC1. The connective [however], on the other hand, is underused in both LC and RC1 as is [thus].

3.2.1.1 The ten most frequently used connectives
On the basis of the quantitative analysis of the learner corpus the following ten connectives are used most frequently.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Connective</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[because]</td>
<td>41</td>
</tr>
<tr>
<td>2</td>
<td>[also]</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>[while]</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>[still]</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>[again], [example]</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>[however]</td>
<td>9</td>
</tr>
</tbody>
</table>
3.2.1.2 Misuse of connectives in learner corpora

On the basis of the ten most often used connectives the frequency of misuse is established. In table 3 the frequency of use is related to the frequency of misuse providing the percentage of misuse.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Connective</th>
<th>Frequency of use</th>
<th>Frequency of misuse</th>
<th>Percentage of misuse</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[because]</td>
<td>41</td>
<td>13</td>
<td>31%</td>
</tr>
<tr>
<td>2</td>
<td>[also]</td>
<td>28</td>
<td>5</td>
<td>12%</td>
</tr>
<tr>
<td>3</td>
<td>[while]</td>
<td>14</td>
<td>2</td>
<td>14%</td>
</tr>
<tr>
<td>4</td>
<td>[still]</td>
<td>13</td>
<td>5</td>
<td>38%</td>
</tr>
<tr>
<td>5</td>
<td>[again], [example]</td>
<td>11, 11</td>
<td>0, 4</td>
<td>0, 36%</td>
</tr>
<tr>
<td>6</td>
<td>[however]</td>
<td>9</td>
<td>5</td>
<td>56%</td>
</tr>
<tr>
<td>7</td>
<td>[finally]</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>[in fact], [therefore], [thus]</td>
<td>4, 4, 4</td>
<td>0, 1</td>
<td>0, 25%</td>
</tr>
<tr>
<td>9</td>
<td>[instead]</td>
<td>3</td>
<td>2</td>
<td>67%</td>
</tr>
<tr>
<td>10</td>
<td>[conclusion], [obviously]</td>
<td>2, 2</td>
<td>0, 0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3: Frequency of use and misuse and percentage of misuse of most frequently used connectives in LC.

Table 4 contains a ranking of the most often misused connectives based on the percentages provided in table 3.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Connective</th>
<th>Percentage of misuse</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[instead]</td>
<td>67%</td>
</tr>
<tr>
<td>2</td>
<td>[however]</td>
<td>56%</td>
</tr>
<tr>
<td>3</td>
<td>[still]</td>
<td>38%</td>
</tr>
<tr>
<td>4</td>
<td>[example]</td>
<td>36%</td>
</tr>
</tbody>
</table>

Table 4: The ten most frequently used connectives in LC.
The five most frequently misused connectives in table 4 have semantic, syntactic or stylistic properties on the basis of which they are misused by language users. Sometimes there is a combination of misuse with regard to properties in which case each example of misuse is included. Interestingly, there is little correlation between the number of times a connective is used and the corresponding percentage of misuse. For example, the connective [instead] is only used twice in the LC but its 67% misuse scores the highest percentages of all connectives in this analysis. The more frequently used connective [because], on the other hand, has the lowest percentage of misuse.

Table 5 shows how often a connective is misused on the basis of each of the three properties. It is clear that stylistic misuse is the most prominent. Below are examples of the misused connectives and the different types of misuse. All data are drawn from the learner corpus. For each example connective the context sentences are provided along with the qualitative analysis of the problem and suggested alternatives. More information on the correct use of connectives can be found in the connectives chart (see Appendix 1) and in the design research of this project.

Examples of misuse:

1. [instead] (stylistic misuse)

Context sentence(s):
[It makes him realize how selfish he is until they show him. When the police come, his friends leave Alex behind, **instead** of helping him. So Alex gets caught by the police.]

**Problem analysis:**
The problem here is caused by the placing of the comma in front of [instead] making it an example of stylistic misuse. One could also argue that the faulty use of commas results in obscured syntax. As we will see, the use of commas in English in general and particularly in relation to connectives is particularly problematic to learners.

**Correct alternative:**
Omit the comma in front of [instead]: [... leave Alex behind **instead** of helping him.]

2  **[however]** (syntactic and stylistic misuse)

**Context sentence(s):**
[Some of Frodo’s friends are skilled fighters and they win most of the time. One of his friends is **however** killed by orcs. This only enlarges Frodo’s will to defeat the enemy and he continues his mission.]

**Problem analysis:**
This is a clear example of faulty syntax where the connective is placed behind the main verb instead of in front of it. In addition, the commas used to indicate a pause are omitted.

**Correct alternative:**
The connective should be placed in front of the verb and in between commas: [... of his friends, **however**, is killed by orcs.]

3A  **[still]** (stylistic misuse)

**Context sentence(s):**
[He zapped, for example, an old man who had drunk from his glass of water. He was disgusted and the man did not even apologise. Worse **still** the man had been eating poached eggs.]

**Problem analysis:**
The learner uses the additive connective [still] or, in this specific context, collocational connective [worse still], semantically correct but forgets to include the ‘pausing’ comma.

**Correct alternative:**
A comma should be inserted behind the connective collocation: [Worse **still**, the man had been eating poached eggs.]
3B  [still] (syntactic misuse)

Context sentence(s):
[His time in jail made him realize more and more what he had done and that this wasn’t the right thing. Though still he was committing violence in jail as well. Until he got picked for the Ludovico technique]

Problem analysis:
The connective is placed in the wrong order presumably caused by mother-tongue interference. In addition, the linking of the first and second sentence is incorrect.
Correct alternative:
Place connective in between first en second verb: [Though he was still committing violence in jail as well.] In addition, the connective [though] is to be replaced by [However,] or the two sentences should be linked.

4  [example] (syntactic and stylistic misuse)

Context sentence(s):
[There is not really an antagonist. There are several suspenses in the story, for example the title; you don’t find out immediately who The Thunderbolt Kid actually is.]

Problem analysis:
The problem described here is a common one. The use of the connective [for example] in front of the subject in this run-on sentences is not correct.

Correct alternative:
The sentences should be separated and the following sentences should be rewritten: [There are several suspenses in the story. For example, the title does not immediately give away ...]

5  [because] (stylistic misuse)

Context sentence(s):
[Bobbie cut herself so she would bleed, which proved that she was no robot. The story had an open end, because the book didn’t mention what happened at Bobbie’s place when she cut herself.]

Problem analysis:
The connective [because] is preceded by a ‘pausing’ comma.

Correct alternative:
The comma in front of [because] should always be left out: [The story had an open end because the book didn’t mention what happened at Bobbie’s place when she cut herself.]

3.2.1.3 The most obstructive connectives and misuse
The various findings described above automatically lead to the answer to the final research question at this stage of this investigation: which connectives and types of misuse are most obstructive in creating appropriate connectivity and which, consequently, require most attention in (remedial) teaching? The quantitative analysis of the learner corpus has shown that the five most frequently misused connectives are (in order of frequency): [instead], [however], [still], [example] and [because]. The analysis of the types of misuse of connectives (based on semantic, syntactic and stylistic properties) conveys that stylistic misuse is most prominent, followed by syntactic misuse. Contrary to expectation, linguistic aspects – punctuation and syntax in particular – are more obstructive in establishing appropriate connectivity than conceptual or cognitive properties are.

With a view to remedial teaching, which will be the focus of the design research, it seems logical to use the five connectives described above as a starting point for the design of remedial activities and strategies. In addition, the mechanics of punctuation and word order, the latter of which is not necessarily a connectivity-related field of language use, require close attention.

3.3 Evaluation
This chapter looks at the conclusions that arise from the quantitative and qualitative analyses of the learner corpus in relation to the reference corpora. These will be accompanied by a critical review of the analytical research to determine to what extent the current findings are reliable and therefore transferable to other research settings. Finally, recommendations will be made.

3.3.1 Conclusions
The quantitative analysis of the corpora conveys a definite correspondence between the use of connectives in the learner corpus and the two reference corpora. This consistency in usage is to some extent surprising as it was expected the learner corpus would convey a more significant degree of underuse. However, usage of a number of connectives conforms to initial expectations. For example, the connectives [again] and [because] are overused in
both LC and RC1. The connective [however], on the other hand, is underused in both LC and RC1 as is [thus]. Though it is beyond the scope of this study to provide plausible explanations for this observation, the discrepancy may arise from the fact that LC and RC1 were written by secondary-school students rather than the RC2 university students. This academic characteristic, in turn, may be accounted for by two features: the RC2 students have had more practice in academic writing and/or LC and RC1 have not yet developed the conceptual or cognitive skills to effectively use specific connectives.

With a view to future studies of this sort, some critical notes must be made on the size, nature and use of the three different corpora. First of all, there is a significant number of zero value results in relation to the total number of results. Two reasons for this large number of zero values may be posed here. Firstly, students have too small a range of connectives at their disposal and have not used the zero value connectives, in which case the focus should turn to underuse of connectives. The second and more likely reason is concerned with the overall size of the learner corpus which may be too limited to produce reliable and effective results.

A second critical note must be made about the pieces of student writing constituting the corpora. These may be comparable up to a certain extent but can also contain crucial differences between the linguistic and/or cognitive abilities of the groups compared thus affecting the required outcomes at a linguistic level.

Thirdly, it is highly recommended that the quality and reliability of all written input is accounted for using standardised corpus and learner profiles. Web-based materials, such as RC1/123freeme used in this study, often lack clear profiling and can make results unreliable.

Fourthly, the limited size of the corpus is the logical result of the limited number of participants. This limited number of participants may, in turn, affect the reliability of the quantitative outcome. For example, if one student out of the total of 25 finds the use of one connective problematic the overall percentages of misuse by this group is influenced disproportionately making results unreliable. Another point of attention in this context is the relationship between the sentence length and the corresponding frequency and use of connectives. It is assumed that the production of short(er) sentences increases the use of connectives as learners (and readers) have fewer implicit means to include (or deduce)

---

59 An example of such a learner profile can be found in appendix 2 and at the International Corpus of Learner English at the Louvain Centre for English Corpus Linguistics. See http://cecl.fltr.ucl.ac.be/Cecl-Projects/Icle/icle.htm#request.
syntax-generated cohesion and will therefore be inclined to compensate through extra use of lexical item i.e. connectives. In conclusion, it is recommended to compile much larger learner corpora, contrasting those against comparable reference corpora the reliability of which can be accounted for on the basis of corpus and learner profiles.\textsuperscript{60}

The number and nature of connectives used in the quantitative analysis also require attention at this stage. Due to restrictions of scope of this study the number of extracted connectives has been limited to thirty. However, to be able to provide a more comprehensive picture of connectivity usage by learners a greater number of connectives would have to be included, permitting these connectives are clearly defined and classified and used in a univocal manner. Furthermore, computer-generated concordances do, as yet, not distinguish between the various forms and functions of connectives. As such, concordance searches on [for] and [as], for example, yield not only the required connectives but all other forms and functions as well making manual revision and selection absolutely essential, even with extensive corpora.

The next stage of the analysis is concerned with the extent and nature of the misuse of connectives. The quantitative results here are used to determine which five connectives are misused most frequently. This, again, may be the result of the restricted size of the learner corpus. The limited yet inappropriate and therefore high-ranked misuse of the connective [instead] may be linked to one learner thus obscuring its validity in more general terms.

The qualitative analysis reveals the type of inappropriateness of each of the five most frequently used connectives. The appropriateness may be the result of the obscured use of semantic, syntactic or stylistic properties of the connective. Contrary to expectations, stylistic misuse is by far the most common problem, followed by syntactical inappropriateness. This may indicate that students have fewer problems understanding and using the underlying conceptual ideas that these connectives denote than initially assumed. It may, however, also simply be that students are unable to use of a greater range of connectives, as stated before. Though not formally a part of this study, the significance of underuse of connectives should also be emphasized in this context. Underuse of connectives, i.e. the absence of connectives and the subsequent lack of connectivity, can only be established through a qualitative analysis of corpora whereby results have to be produced entirely manually. This obviously

\textsuperscript{60} In compiling a learner corpora it is important to bear in mind at which point(s) in the learners’ language development data is collected and how this moment may influence the results and conclusion in comparison to the reference corpora. For example, one may want to diagnose the use of connectives by different learner groups at a specific moment or follow their development over a certain time span. In the latter case, it is important to realise that the learners’ usage of connectives will have progressed and that different levels of proficiency in connectivity usage are combined in the same corpus.
requires a more exhaustive approach to the analysis of the corpora. On a final note it was found that the usage regularly conveyed inappropriateness that was the result of combined properties e.g. syntactic and stylistic. This, then, should be the first starting points in the design of remedial materials.

3.3.2 Recommendations

It can thus be concluded that this study should be considered a first and modest step in yielding results about the usage of connectivity by Dutch adolescent learners of English at advanced level. In a next stage it is essential that a more extensive and therefore more representative learner corpus is compiled, consulted and contrasted against the BAWE corpus. This automatically implies that a more elaborate list of connectives is to be used. With a view to designing remedial materials this analysis shows that stylistic and syntactic properties should be prioritised (but certainly not isolated from) semantic ones allowing learners to understand and use connectives through corpus-based reference materials.
4.0 Introduction

This chapter presents the second part of this study i.e. the design research. Using some of the conclusions and recommendations in the analytical research as its starting points, chapter 4 revolves around a corpus-based classroom intervention. The aim of this experiment is to find out if and to what extent improvements in connectivity usage among the target group can be achieved. Before looking at the nature and scope of the intervention and the research context in which it is embedded, its relationship with the analytical research and its methodological starting points are briefly explored. The broader theoretical framework of connectivity and corpus-based language teaching is presented in chapter 2.

4.0.1 Link with analytical research

The outcome of the quantitative analysis of the learner corpus conducted in the analytical research of this study, yielded no significant difference in the number of connectives used by the Dutch TESOL learners as compared to the reference corpus. In addition, the qualitative analysis of the LC conveyed a relatively high level of appropriate use of connectivity with regard to stylistic, syntactic and semantic properties. Contrary to expectations, however, stylistic and syntactic inappropriateness occurred more frequently than semantic misuse. In those instances where inappropriate use was diagnosed it was often the result of combined properties e.g. syntactic and stylistic ones. From these findings, then, it may be concluded that the target group of this study, and Dutch TESOL learners in general, perhaps, do not have significant problems with connectivity when compared to TESOL learners with different native-language backgrounds. After all, English and Dutch share many common linguistic grounds and English has ‘infiltrated’ Dutch culture and language more so than it has other languages. However, it was also said that the results the analysis yielded may not have been entirely representative due to the limited size of the learner corpus, the number of connectives used and the exclusion of the analysis of underuse of connectives. In addition, statistical research carried out in the 1980s has repeatedly shown there to be no difference between the number of connectives used in good and poor native-American English essays. It was found that the focus should be on how connectives are used to achieve cohesion and coherence, implying a stronger focus on the qualitative analysis of texts.\(^{61}\)

\(^{61}\) Granger & Tyson 1996: 17.
In chapter 1.1 we have seen this claim substantiated by other authorities in the field along with a plea for alternative teaching methods and materials. Furthermore, a small-scale survey among TESOL teachers in the Netherlands was conducted in the scope of this study. The outcome conveys that a significant number of teachers consider connectivity a ‘problem area’ to teach, which was the assumption for the analytical research study.

The design research, therefore, takes as its starting point that the appropriate use of connectives in formal writing still poses a worthwhile challenge to Dutch (adolescent) learners of English. It is assumed that this specific area of English foreign-language teaching, then, requires intensive and effective remedial attention and practice if we are to upgrade the students’ performances to the desired CALP level. In line with the conclusions and recommendations of the analysis, the emphasis will be on stylistic and syntactic features, as these were found to be most problematic, and on an integrated approach. To facilitate these ambitions and to be able to determine differences in effectiveness in comparison to more conventional approaches to connectivity acquisition, the intervention will make use of practical applications of COCA and its tools, and tailor-made exercises and tasks. The ultimate aim, of course, is to find out if this type of intervention yields a higher degree of appropriate connectivity usage than conventional approaches do.

4.1 Procedure
The design research combines recent insights in connectivity teaching and corpus linguistics with an experimental classroom intervention. This chapter introduces the research procedures and the intervention.

4.1.1 Hypothesis, research question and conceptual model
The hypothesis of this study is that a corpus-driven approach to connectivity acquisition and teaching is more effective than more conventional approaches are. Effectiveness in this context is defined in quantitative and qualitative terms. Quantitatively, it is expected that the increase of appropriate connectivity usage among the experiment group will be statistically more significant than that of the control group because of the effects of the intervention. Qualitatively, the results of the experiment will provide insight into which corpus and connectivity aspects of the intervention the learners found useful.

63 The survey was conducted using the mailing list of the English community at mijn.digischool and required teachers to comment on whether or not they encountered problems in teaching and assessing connectivity. All respondents said they found it problematic to some extent.
The primary research question can be formulated as:

*Is a corpus-driven approach to the acquisition of connectives among advanced adolescent learners of English in the Netherlands more effective than more conventional approaches are?*

**Secondary research question 1** (quantitative):

*Does a corpus-driven approach to the acquisition of connectives yield a statistically significantly higher level of appropriate syntactical, semantic and stylistic usage of connectives than conventional approaches do?*

**Secondary research question 2** (qualitative):

*Does the experiment group perceive a corpus-driven approach to the acquisition of connectivity useful?*

In order to identify any increased effectiveness of connectivity skills the results of the experiment group will be tested against those of the control group. The experiment group will be exposed to the intervention; the control group will follow a more conventional approach to connectivity teaching. The increased level of appropriate connectivity usage is based on three linguistic properties: syntactical, semantic and stylistic properties. To find out if and how the experiment group benefited from the intervention they will be asked to fill out a short questionnaire containing open questions. The answers will be used to find out about students’ perception of the intervention.

**Variables and conceptual research model**

The variable in question here is the degree and nature of the learning effect that the intervention will have on appropriate usage of connectives with respect to their semantic, syntactic and stylistic properties. It is expected that after the intervention students will use connectives more effectively than before the intervention and that will find a corpus-based approach to connectivity acquisition useful. From this assumption the following conceptual model follows:
4.1.2 Research plan

The core of the design research is the evidence-based intervention i.e. the classroom application used to test if and in what way the intervention has been successful. The plan below describes the types of research and the proceedings of the design research and the relationship between the intervention and the research.

4.1.2.1 Types of research and triangulation

The main research question consists of two different types of research questions. Minor research question 1 (quantitative research) is answered using the numeral difference between the samples at pre- and post-intervention stages of the research. The variable in question is investigated on a ratio scale using the T-test to find out whether the effect size is statistically valid or not. Minor research question 2 (qualitative research) is answered through inquiry research consisting of a questionnaire and an inventory of personal responses.

As to triangulation, the quantitative and qualitative results are expected to reinforce one another i.e. the greater the validity of the effect size of the intervention the greater the appreciation of the intervention, using the outcome of the questionnaire as the research instrument (see chapter 4.1.3 below for details).

4.1.2.2 Proceedings of the research

Participants
The participants consist of fifty Dutch secondary-school students aged 16 or 17 and currently in their fifth year of TVWO. They are generally motivated for English and school in general and have followed a four-year course in bilingual education taking six subjects taught primarily in English. They will finish their English course in the International Baccalaureate Diploma Programma English A2 Higher Level. The students’ general writing proficiency at the
CEFR would at this stage be assessed at B2/C1 level. Though learners have received explicit previous training in connectivity, it is assumed that the majority of the students find linguistic as well as cognitive aspects of connectivity and cohesion in general in both TL2 and TL1 demanding, as has been discussed before. Cognitive limitations specific for this age group and inevitable inappropriateness in connectivity should certainly be taken into account in any analysis of connectivity usage and design of materials.

**Description of research plan**

All students participate in a set module on comparative textual analyses in which they learn to analyse pairs of written texts and write effective comparative commentaries on them. As part of this module, learners will be exposed to connectivity and its relevance in advanced usage of English (see chapter 4.1.4 for details).

All learners do a pre-test requiring them to apply their connectivity skills. The exercises each focus on semantic, syntactic and/or stylistic properties. The outcome of the pre-test is the baseline measurement used to determine the degree of the effects after the intervention has been carried out.

The students are then placed into two groups of 25 students. **Group A** is the control group and will not be exposed to the intervention. Instead, it will do conventional activities on connectivity taken from existing materials such as textbooks and web sources. **Group B** is the experiment group and will be exposed to the intervention focussing on the use of corpora and corpus tools.

After the intervention, both groups do the post-test. This test is identical in format to the pre-test and it is analysed and marked in the same way as the pre-test is. The experiment group will also answer a limited set of reflective questions requiring them to comment on the value of each of the activities they carried out during the intervention. Their ranking and comments will provide insight into how useful and effective they thought the intervention was.

Finally, the results of the pre-test (zero measurement) and the post-test for all learners are analysed quantitatively. In addition, the outcome of the reflective questions is processed for qualitative purposes. The quantitative results are then used to determine whether the intervention has been effective.
4.1.3 Research instruments

The design research makes use of two research instruments, one for the yielding of quantitative data and one for qualitative:

1) Pre and post-tests for quantitative analysis
The pre and post-tests generate scores for the control and experiment groups are used to statistically determine if the intervention as a whole has been effective or not. There are also subcategories for the assessment of semantic, syntactic and stylistic scores. See chapter 4.2.1 (data collection) for a more detailed description of this research instrument. See appendix 3 for the tests and the keys.

2) Questionnaire for qualitative analysis (for experiment group only)
In addition to quantitative data the design research also aims at finding out what students found useful about the intervention. To this end, students in the experiment group will do a questionnaire consisting of ten statements and five open-ended questions inviting them to rank their perception of its usefulness. See chapter 4.2.1 (data collection) for a more detailed description of this research instrument. See appendix 4 for the questionnaire.

4.1.4 Description of intervention

As stated above, the intervention is carried out in the framework of a module on comparative textual analysis that all students (control and experiment groups) take. In this respect, the intervention is part of a greater and authentic learning context. This chapter briefly describes the intervention as product and process reporting on such aspects as teacher role, contents and aims.

4.1.4.1 Methodological starting points and framework
An essential methodological starting point for the effective workings of the intervention is that of the independent language learner. Rather than providing learners with guided face-to-face classroom instruction, teachers are to focus on how to make learners aware of the self-instructive merits of corpus-based queries. Encouraging learners to develop linguistics systems of their own is the basic idea behind ‘data-driven (language) learning’, as discussed in chapter 2.2.2.
However, bearing the 'adolescent needs' of the age group in mind this is easier said than done. It is pivotal, therefore, that learners are exposed to the direct benefits of corpus-based queries at all stages and that parallels are drawn with other supportive materials such as (digital) dictionaries and reference texts. Corpora are particularly useful for (written) revision purposes so it is essential that students learn to use corpora beyond the walls of the classroom. In this way, the corpus can become a useful type of tool for further language development.

The aims of the intervention are for learners to improve their use of connectives; to become aware of the self-instructive nature and merits of corpus-driven language learning; and to become (more) enthusiastic for the potential of corpora for language learning.

4.1.4.2 Intervention as product
The intervention consists of a series of self-instructive lessons. The materials are distributed by means of a printed handout. Samples of class materials and activities can be found in appendix 5. After a plenary session in which the basic aims and tools are introduced students work independently and carry out a number of progressing activities. Individual teacher supervision is available but kept to a minimum as the ultimate aim of corpus-driven language acquisition is for students to be self-supportive.

4.1.4.3 Intervention as process

*Session 1 (45 mins): Introduction to (applied) corpus linguistics and connectivity*
This introductory session explores the theoretical premises of corpus linguistics and its relevance to language and, more specifically, connectivity development. Examples and characteristics of corpora will be shown and the workings of one accessible and large-scale online corpus, COCA, will be introduced. The aim of this session is for students to have a good idea of the workings and relevance of applied corpus linguistics and corpora in relation to connectivity skills.

*Session 2 (45 mins): Exploration of COCA and corpus analysis*
This is a hands-on session in which students learn to work with the basic corpus tools of COCA by carrying out some general activities. The aim of this session is for students to be able to find their way within COCA.
Session 3 (90 mins): Using COCA to explore and improve connectivity skills
In this session learners use the tools introduced in steps 1 and 2 to identify, remedy and improve their own connectivity skills. Through guided activities of a limited scope and focussing on appropriate connectivity usage students will learn how to use the COCA corpus for connectivity purposes in general and to remedy their own problems more specifically. The focus of the activities changes from passive/receptive to active/productive as the students go on. The aim of this step is for learners to develop a good view on their own usage of connectivity and to know how to remedy any problems using COCA.

Session 4 (45 mins): Using COCA to revise connectivity skills
This concluding session requires student to utilise the COCA potential for revision purposes, ultimately aiming at independent use of COCA to improve connectivity skills. The input consists of the revising of the (marked) pre-test using COCA.

Session 5 (45 minutes): Post-test
In the final session the students in both the control and experiment groups sit the post-test, which is used to determine the effects of the intervention.

4.1.4.4 Validity, reliability and generalisability
The research and the intervention were carried out in order to yield generalisable results. However, a few potential weaknesses specific to the research context in this study have to be addressed. First of all, it should be noted that the participants have a higher than normal knowledge of and experience with connectivity due to the tradition of teaching and syllabus at the school. It is expected, therefore, that control and experiment groups in alternative settings would receive lower general scores for appropriate connectivity usage. Contrary to this bias, it is not likely that the pre-test (zero value) has affected the outcome in any way as participants had experience with connectivity before and were used to carrying out the type of assignments. As to the interaction of history and treatment of the research, one should bear in mind that progress in classroom applications of corpus is made - albeit very gradually - and that the current learners’ (and teachers’) relative inexperience with them and the deviating effects this may have had, will be less intrusive in the future.
4.2 Implementation
This chapter describes the actual implementation of the design research i.e. the collecting of the data for the quantitative and qualitative analyses, and the corresponding results.

4.2.1 Data collection
For the data collection also see the research instruments described in chapter 4.1.3.

Quantitative data collection
The control and experiment groups, consisting of 25 learners each, will do a pre-test and post-test before and after the intervention. The two tests comprise exercises of similar levels, templates, organisation and duration. Students can score a total of 29 points which will be equated to a scale of 1 for 10 for statistical output purposes. The tests consist of three parts with a number of exercises, each focussing on semantic, syntactic and stylistic features of connectives. The respective maximum scores for semantics, syntax and stylistics are 17, 8 and 4. Within each part there is a gradual shift from receptive to (semi-)productive exercises. The scores for the respective parts will also be used to identify which specific property of connectives students have made most progress with. In this way, the effectiveness of each of the three different areas can also be assessed. The scores of the pre-test and post-test of the control and experiment group, will be statistically tested to see whether or not the intervention has been effective or not. Though not related to the main research question, students are also required to write down the time (in minutes) they spend on each of the two tests. This will shed light on any increased speed with which the students complete the task and can be used for further research purposes. Please see appendix 3 for the pre- and post tests and the keys.

Qualitative data collection
In addition to quantitative data used to determine effectiveness the design research also aims at finding out what students found useful about the intervention or what they see fit for improvement. To this end, students in the experiment group will do a questionnaire consisting of ten statements and five open-ended questions. The ten statements aim at finding out how useful connectivity and corpus analysis were perceived by the students by ranking them on a scale of one to five. The first five statements focus on connectivity and its three areas. The latter five statements invite students to rank the usefulness of corpus linguistics and COCA.
The five open-ended questions allow students to give feedback on which aspects they found interesting, demanding and worth exploring with. It also asks students to provide practical recommendations. Please see appendix 4 for the complete questionnaire.

4.2.2 Results
Below the quantitative and qualitative results are presented.

4.2.2.1 Quantitative results
The statistical output and the tables were generated using the statistical package SPSS. The tables below show the paired statistical output of the average scores of the pre- and post-tests of the control and experiment groups.

Table 1 below shows that the average marks (mean) of the pre-test of both the control and experiment groups are very similar. The standard deviation is small. This was expected given the common starting points and relative homogeneity of the control and experiment groups.

<table>
<thead>
<tr>
<th>Paired Samples Statistics</th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average mark pre-test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>control group</td>
<td>6,2640</td>
<td>25</td>
<td>.79366</td>
</tr>
<tr>
<td>Average mark pre-test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>experiment group</td>
<td>6,2920</td>
<td>25</td>
<td>.84800</td>
</tr>
</tbody>
</table>

Table 1: Paired sample statistical output of pre-test of control and experiment group.

Table 2 below shows the difference in average marks of the post-tests of the control and experiment groups. The statistical difference is significant.

<table>
<thead>
<tr>
<th>Paired Samples Test</th>
<th>Paired Differences</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Deviation</td>
<td>Std. Error</td>
<td>95% Confidence Interval of the Difference</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>Std. Error Mean</td>
<td>Std. Error Mean</td>
<td>Lower</td>
</tr>
<tr>
<td>Pair 1</td>
<td>Average mark post-</td>
<td>-1,02400</td>
<td>.07234</td>
<td>-1,05386</td>
</tr>
<tr>
<td></td>
<td>test control group</td>
<td>- Average mark post-</td>
<td>test experiment group</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Paired sample T-test showing the mean and statistical significance.
4.2.2.2 Qualitative results

Ranking of students’ statements

The tables below show the students’ rankings and answers on the basis of the questionnaire. The average ranks are set against the following statements: 5 = 'I agree completely.'; 4 = 'I agree for the greater part.'; 3 = 'I agree to some degree.'; 2 = 'I hardly agree.'; 1 = 'I absolutely do not agree.'

Table 3 below shows the ranking of statements based on students' perception of the usefulness of connectivity. The average ranking of the statements on usefulness of connectivity is 3.5 equalling the statement 'I agree for the greater part/to some degree.' Statement 5 has a ranking of 2.8, which is significantly lower than the others: Students only agree 'to some degree' about increased confidence about connectivity usage after the intervention.

<table>
<thead>
<tr>
<th>Statements in relation to connectivity</th>
<th>Average student ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Aware of the importance of connectivity</td>
<td>3.7 ('I agree for the greater part.')</td>
</tr>
<tr>
<td>2) Usefulness of semantic exercises</td>
<td>3.2 ('I agree to some degree.')</td>
</tr>
<tr>
<td>3) Usefulness of syntactic exercises</td>
<td>3.8 ('I agree for the greater part.')</td>
</tr>
<tr>
<td>4) Usefulness of stylistic exercises</td>
<td>3.9 ('I agree for the greater part.')</td>
</tr>
<tr>
<td>5) Increased confidence about connectives</td>
<td>2.8 ('I agree to some degree.')</td>
</tr>
</tbody>
</table>

Table 3: Ranking of statements based on perceived usefulness of connectivity.

Table 4 below shows the ranking of statements based on students' perception of the usefulness of corpus tools and COCA. The average ranking of the statements on usefulness of corpus tools and COCA is 3.2. This rank equals the statement 'I agree to some degree.' Statement 3 has a ranking of 2.7, which is relatively lower than the others: Students only agree 'to some degree' about the future use of COCA after the intervention.

<table>
<thead>
<tr>
<th>Statements in relation to corpus tools and COCA</th>
<th>Average student ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Corpora are useful tools</td>
<td>3.3 ('I agree to some degree.')</td>
</tr>
<tr>
<td>2) Corpora are more useful than dictionaries.</td>
<td>3.2 ('I agree to some degree.')</td>
</tr>
<tr>
<td>3) Predicted future use of COCA</td>
<td>2.7 ('I agree to some degree.')</td>
</tr>
</tbody>
</table>
### Inventory of students’ responses

Table 5 below is an inventory containing answers to the open questions produced by each of the 25 students in the experiment group. For each question it gives the number of times student provided similar responses. For questions 1 to 4 only answers which appeared three or more times have been recorded. Question 5 was entirely open inviting students to provide suggestions for improvement. Not every student provided answers to each question.

<table>
<thead>
<tr>
<th>Focal ideas in question and student response</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1) The greatest benefits of corpora:</strong></td>
<td></td>
</tr>
<tr>
<td>Locating large numbers of examples of words</td>
<td>16</td>
</tr>
<tr>
<td>Providing large numbers of examples in good English</td>
<td>4</td>
</tr>
<tr>
<td><strong>2) The most interesting aspects of corpora:</strong></td>
<td></td>
</tr>
<tr>
<td>Realising the size of corpora</td>
<td>10</td>
</tr>
<tr>
<td>Providing different contexts/meanings</td>
<td>6</td>
</tr>
<tr>
<td><strong>3) The most difficult aspects of corpora:</strong></td>
<td></td>
</tr>
<tr>
<td>Choosing a select set of relevant examples from millions of examples</td>
<td>4</td>
</tr>
<tr>
<td>Finding the specific meaning of a word on the basis of context</td>
<td>14</td>
</tr>
<tr>
<td>Checking stylistic features of connectives</td>
<td>2</td>
</tr>
<tr>
<td><strong>4) The greatest disadvantage of corpora:</strong></td>
<td></td>
</tr>
<tr>
<td>Drawing conclusions yourself which may be wrong</td>
<td>6</td>
</tr>
<tr>
<td>Being presented with too much information</td>
<td>4</td>
</tr>
<tr>
<td>Time consuming to find the correct context</td>
<td>4</td>
</tr>
<tr>
<td><strong>5) Further comments or suggestions? (not filled out by every student)</strong></td>
<td></td>
</tr>
<tr>
<td>Useful but I will never use it at home</td>
<td>2</td>
</tr>
</tbody>
</table>
4.3 Evaluation

In this chapter the results of the quantitative and qualitative analyses are interpreted and evaluated. In addition, recommendations are made for revised interventions, departmental and school policies and further research.

4.3.1 Conclusions

Minor research question 1 set out to discuss whether a corpus-driven approach to the acquisition of connectives yields a higher level of appropriate usage of connectives than conventional approaches do. The quantitative analysis has shown that the effect size of the intervention is in fact statistically significant. As such, the initial hypothesis can be confirmed: the increased effectiveness of connectivity usage among learners in the experiment group as compared to those in the control group can be ascribed to the experimental nature and workings of the intervention.

Minor research question 2 set out to rate the learners' perception of the usefulness of a corpus-driven approach to the acquisition of connectivity. The outcome of the questionnaire, by nature of qualitative research, shows a more varied picture. Learners rated connectivity acquisition in general as useful 'for the greater part/to some degree'. Of the three linguistic properties of connectives, semantic properties were considered to be least useful. Interestingly, students rated their increased confidence in connectivity skills 'to some degree', which is a significantly lower rating. A possible answer to this outcome may be the fact that semantic properties are often considered to be the most demanding - or most elusive - aspect of appropriate connectivity acquisition, which is reconfirmed by the students' observation of having limited confidence with them. In addition, the methodological starting points aimed at self-instruction and exploration (rather than classroom explanation) and the students' relative inexperience with this approach may also account for their limited confidence.
The usefulness of corpus tools and COCA was rated by learners as useful 'to some degree.' Though the average outcome is neutral, the statement asking students about their use of COCA after the intervention receives the relatively lowest rating. At this stage no satisfactory explanation for this rating can be given but it is worthwhile to observe that there seems to be a correlation between the learners' (relative) 'lack of confidence in connectivity usage' and (relative) 'little future use' of corpora.

An interesting contradiction in the inventory of student responses is the observation that one of greatest benefits of corpora is their size and the ability to locate large numbers of examples in corpora whilst, at the same time, they consider one of the more daunting aspects of corpora the locating of a specific meaning of a word on the basis of context.

On the basis of the above, it can be concluded that the intervention as a product aiming at improved connectivity skills for the experiment group has been successful. The effects of the intervention as a process, however, are less convincing as learners still feel relatively ill-equipped to deal with large quantities of input and make satisfactory choices.

4.3.2 Recommendations for the English department and school

The recommendations below focus on the improved design and implementation of future interventions, on school and departmental policies and on further research in the fields of connectivity and corpus-based TESOL instruction.

The linguistic definitions of connectives in both the analytical and design research of this study and in the intervention require more consistent specification as do the subsequent results yielded. For syntactical properties, comma usage at this stage needs to be more specific (or abandoned) as there is much clarification about appropriate usage is left to be desired, even when established corpora (such as COCA) are used. For stylistics, clearer instruction about register and formal and informal language use is required. It is strongly recommended this aspect is integrated into a broader and more profound familiarisation with formal and academic language production. The use of specialised or differentiated corpora are very suited for this purpose. Also, rather than focussing on all three linguistic properties of connectives at the same time, a future intervention may be centred around just one of them. Alternatively, the instructor may have learners explore a select number of connectives in depth, though this limitation in exploration would affect the self-instructive nature of corpus explorations. Such selections may be based on the findings of analytical research as has been done in this study. By delving into issues in depth rather than in breadth quantitative
and qualitative results in the form of learner observations and learner problems are easier to identify and remedy. Consistency is also important in the formulation, weighting and marking of the exercises in the pre- and post tests: these need to have the same nature and degree of difficulty.

To increase the statistical validity of the results certain practical measures can be taken at departmental and school level. Firstly, as is appropriate to educational research at post-graduate level, this study aimed at making conclusions beyond the individual classroom i.e. at departmental and school level. However, the fact that all participants were trained by the same instructor is likely to have affected the outcome. Therefore, it is essential that at a next stage (relevant) learner groups taught by different teachers and with varying backgrounds are involved. Secondly, it is essential that enough class time is reserved for the intervention. To make sure that certain groups do not lag behind it may be advisable to have all students in a certain form participate in the intervention. Linked to this, is the ethical problem of the 'neglected' position of the control groups who do not participate in the intervention and thus do not benefit from any selective language advancing. A suggestion would be to do the same intervention twice: once strictly for research purposes involving only the experiment groups and once for classroom purposes, during which learners from the experiment group may supervise learners from the control group. Of course, during the actual intervention it is important that the control group develops a sense of being taken seriously. The additional advantage of this procedure is that triangulation can also be established.

4.3.3 Recommendations for further research and classroom application
With respect to further research and classroom applications and interventions a number of recommendations can be made. To yield more results about typically Dutch patterns and errors in connectivity usage a contrastive analysis of individual connectives is recommended, to be carried out at a much more detailed level than has been the case in this study. Springer has shown that Dutch university-level learners of English overuse connectives of cause and reason (such as [because] and [therefore]) believing that cause clauses are conceptually easier to produce and that learners suffer from a limited linguistic repertoire.64 Conclusions of this kind can form the basis from which purposeful remedial materials can be designed and used. Teacher-researchers can use the corpus-analytical tool TextStat and analytical model

64 Springer 2012: 178.
based on the three basic linguistic features of connectives, both used in this study, for their own classroom-based research in a relatively easy and straightforward way.

Unlike countries like Germany, China and the United Kingdom, applied corpus linguistics - and any application with reference to connectivity usage or acquisition - is still in its infancy in Dutch research and pedagogic contexts. To facilitate the need for relevant analytical and design-based data a much greater number of corpus-based interventions should be carried out in various learner contexts and at various levels. The findings of these investigations could lead to the development of corpus tools and software - of which the United-States based COCA is a very generic example - geared to the specific needs of Dutch learners of English.

Most important, however, is a reorientation on the pros and cons of prescriptive and descriptive language teaching with respect to the digital age and, more specifically, the self-instructive potential of corpora. More research is likely to show that the merits of corpora in contexts such as CLIL and CBLL are extensive. In this way, self-supported or data-driven language learning based on authentic language generated digitally at the learner’s disposal can be integrated into subject-specific language contexts common to such methods as CLIL and CBLL. The acquisition of appropriate usage of connectives in academic English discourse is but an example of such self-controlled language development and revision.

On a final note it needs to be mentioned that with the concept of 'digital learning' we have, of course, reached an educational challenge exceeding that of the English-language computer room. As some of the results of the questionnaire in the design research revealed and despite the impression some teachers may have of the 'multitasking Einstein generation', adolescent learners still feel insecure selecting relevant data from large quantities of (digital and corpus-based) information independently and systematically. Assuming that Bod is correct in his prediction that digital information generation and retrieval will increase drastically in the near future the teaching community at large had better starting developing more effective learner tools. 65 It is hoped that this study has shown that corpus-based tools and (language) learning are very suited for this purpose.

---

65 Inaugural address by Rens Bod, 6 October 2011, Universiteit van Amsterdam, http://www.english.uva.nl/news/professorappointments.cfm/3476083D-5913-4663-9E9F02DF06446A5A
CHAPTER 5: CRITICAL REFLECTION

This final chapter is a critical and, to a certain extent, personal reflection on the proceedings and the outcome of the research and the intervention. Some of the ideas presented here are also discussed from a somewhat different perspective in chapter 4.4.

5.1 Proceedings and reception of research

Nearly all teachers will acknowledge that professional development is an essential and usually stimulating part of their careers. In the Netherlands in recent years, the more or less traditional role of ‘teacher as educator’ has been supplemented by that of the ‘teacher as educational researcher’, the aim being that teachers are to be equipped with the necessary academic tools and attitudes enabling them to identify, analyse and improve aspects of their professional practice in the classroom, department or school environment. In the daily practice of the average teacher, however, it is no easy task to carry out, reflect upon and amend research the way it should ideally be done. In that respect, this research report is no different. In the two years during which it was conceived and conducted there have been a series of undesirable yet inevitable pauses, caused by time pressure, unavailability of student input or rigid school-related planning and syllabi. Regrettably, these conditions have regularly led to fragmented, discontinued and - how ironic - incoherent chunks of research periods that constitute this thesis. It is my sincere belief that for any type of reliable research ample time is reserved for systematic implementation and reflection - educational research being, of course, no exception. In-depth classroom and subject-based research seems more suited for this purpose than large(r)-scale or school-based educational research, in my perception.

Related to time restrictions is the fact that departmental and school involvement in the research was kept to a minimum and that the research was only carried out by the writer of this report, as has been accounted for above. One of the reasons for this was the relatively specialised and experimental nature of the research topics but also the aforementioned fact that necessary reading, practice and teaching time for colleagues was simply not available. In the foreseeable future, however, I plan to integrate manageable aspects of corpus-based connectivity acquisition in a module on comparative essay writing for all year five students. At school level, students would benefit immensely from a interdepartmental view on the importance of connectives in writing (and speaking) and the remedying roles of corpora and
data-based learning - as they would from any type of useful transferable knowledge and skills.

5.2 Effects and generative potential of research

Though relatively little comparable class-room based research has been carried out in the Dutch context, most of the conclusions of this research seem to correspond with those of prior studies carried out in Germany, China and Spain. First of all, there is a need for more large-scale empirically-based studies to obtain a more accurate view of how connectives \textit{should be} used and \textit{are} usage.\textsuperscript{66} At the same time, these large-scale studies should analyse on the use of individual connectives from a contrastive perspective, as has been stated before, the results of which may then be used for pedagogic purposes.\textsuperscript{67} As we have seen, this study has tried to yield comparable output employing free corpus tools and a specific analytical model. Along with the class materials developed for the intervention, any (language) teacher can implement research along the same lines. Secondly, the research findings reconfirm learners' relative lack of experience - or willingness - with digital corpora as language-supporting and developing storehouses and tools. Marks's observation that many teachers and learners still have 'a preference [...] for straightforward right-or-wrong answers' means that many ways yet need to be paved before the potential of both corpora and data-driven language learning can reach the TESOL classroom.\textsuperscript{68}

\textsuperscript{66} See Granger and Tyson 1996: 18-19.
\textsuperscript{67} See Springer 2012: 178.
\textsuperscript{68} Marks 2010.
TO CONNECT OR NOT TO CONNECT

Corpusgericht analyse- en ontwerponderzoek van Engelse connectieven in teksten van tweetalige VWO-bovenbouwleerlingen

David Geneste

Sleutelwoorden: TESOL; tekst; tweetalig onderwijs; connectieven; corpuslinguïstiek; corpora; data-driven language learning; analyse-onderzoek; ontwerponderzoek; interventie.

Inleiding

Het Europees Referentiekader geeft aan dat leerlingen in de bovenbouw van het VWO op ERK-niveau B1 hun Engelse schrijfvaardigheid moeten beheersen. Van leerlingen die voor het vak Engels het tweejarige 'Internationaal Baccalaureaat' volgen wordt bovendien verwacht dat zij effectieve en samenhangende essays op (pre-)wetenschappelijk niveau leren schrijven. Dezelfde kenmerken en eisen zijn ook van toepassing op het bovenbouwprogramma Engelse taal- en letterkunde van de tweetalige afdeling van het Sint-Maartenscollege te Voorburg, de context waarbinnen dit onderzoek is uitgevoerd.

Een van de voornaamste belemmeringen die taalleerders ondervinden bij tekstproductie is het correct, d.w.z. natuurlijk en vloeiend gebruik van zogenaamde 'connectives' (Nederlands: verbindingen- of voegwoorden). Deze in dit onderzoek gehanteerde (Engelse) verzamelterm duidt op een categorie functionele woorden of woordcombinaties die het onderliggende en vaak abstracte verband tussen andere woorden, zinsdelen, zinnen en alinea's aangeven. Een effectief gebruik van connectives maakt teksten coherenter, logischer en uiteindelijk begrijpelijk voor de lezer. Uit internationaal onderzoek naar textanalyse en -productie blijkt echter dat taalleerders het correct gebruik van connectives inderdaad als lastig ervaren. Een enquête gehouden onder docenten Engels in Nederland lijkt deze constatering voor de Nederlandse context te bevestigen.

Dit praktijkonderzoek heeft het gebruik van Engelse connectives van een specifieke groep Nederlandse TVWO-bovenbouwleerlingen in kaart proberen te brengen. Daarbij is gebruikt gemaakt van recente corpuslinguïstische inzichten en onderzoeksinstrumenten. In de analysefase is gekeken naar de specifieke aard en omvang van het probleem. De hieruit
voortgekomen resultaten zijn gebruikt om het ontwerponderzoek op te baseren, waarbij een experimentele interventie is uitgevoerd om te bezien of deze het gebruik van connectives ook inderdaad kan verbeteren.

_Uitvoering en resultaten van analyseonderzoek_
In het analyseonderzoek is gekeken naar de specifieke eigenschappen van de problemen die leerlingen ondervinden met connectiviteit. De onderzoeksvraag richtte zich op welke connectives leerlingen het meest frequent gebruiken en welke van deze het meest regelmatig incorrect gehanteerd worden. De mate van correct of incorrect gebruik van connectieven is in het onderzoek dus de hoofdvariabele; de criteria voor de aard van de correctheid zijn de semantische (onderliggend verband), syntactische (positie in zin) of stilistische (formeel of informeel karakter) eigenschappen van de connectives in kwestie. De verwachting was dat de doelgroep vooral moeite zou hebben met semantische aspecten gezien de gecompliceerde afwegingen die hierbij gemaakt moeten worden.


De uitkomst van het analyseonderzoek laat zien dat er inderdaad een groep connectieven is dat meer problemen oplevert dan anderen. Er is echter geen direct verband aan te wijzen tussen de aard van incorrect gebruik van de gevonden connectieven. Bovendien is het verschil in correct gebruik van connectieven in de reference en de learner corpora niet statistisch significant. Tot slot bleek - verrassend genoeg - dat de onderzoeksgroep minder problemen te hebben met het semantisch gebruik van connectieven maar juist meer met stilistisch eigenschappen.

_Uitvoering en resultaten van ontwerponderzoek_
In het ontwerponderzoek zijn de resultaten van de analyse in licht aangepaste vorm gebruikt om een experimentele interventie in de vorm van een op corpuslinguïstische inzichten
gebaseerde project, vorm te geven. Centrale vraag hierbij was of het gebruik van digitale corpora en bijbehorende corpustools een hogere mate van correct gebruik van connectieven oplevert dan een meer conventionele aanpak, zoals die bijvoorbeeld in leergangen te vinden is. Ook is middels een vragenlijst getracht te achterhalen hoe de participanten het werken met corpora en corpus tools beleefd hebben.

Om gegevens te genereren werden een voor- en nameting gedaan bij de controle- en experimenteergroep. Tijdens de tussenliggende interventie heeft de controlegroep aan 'conventionele' opdrachten zonder corpora gewerkt terwijl de experimenteergroep werd geconfronteerd met COCA, een publiekelijk toegankelijk corpus, en bijbehorende opdrachten. Als didactisch model is expliciet gekozen voor 'data-driven language learning' waarbij leerlingen gestimuleerd worden om grotendeels zelfstandig uit grote hoeveelheden input (in corpora) relevante keuzes te maken en patronen te herkennen. Deze aanpak vergt een grote mate van zelfstandigheid en keuzevrijheid van de leerlingen en een terughoudende opstelling van de docent.

De kwantitatieve analyse van de onderzoeksgedraggen met behulp van SPSS laat zien dat er een statistisch significant verschil is tussen de prestaties van de controle- en experimenteergroep. In dit opzicht is dus aangetoond dat binnen de hier gehanteerde onderzoekscontext een corpuslinguïstische interventie meer bijdraagt aan een correct gebruik van connectieven dan traditionele methoden.

**Conclusies en aanbevelingen**

Het onderzoek laat een aantal interessante doch deels ook tegenstrijdige uitkomsten zien. Op basis van het analyseonderzoek zou kunnen worden geconcludeerd dat de gesignaleerde connectiviteitsproblematiek voor de onderzoeksgroep niet bijzonder urgent is. Hoewel deze bewering niet aannemelijk is kunnen verklaringen zijn de relatieve ervarenheid met connectieven van de onderzoeksgroep vanuit het bestaande lesprogramma binnen de onderzoekscontext, een te geringe beschikbaarheid van onderzoeksgegevens en participanten, en/of het gebruik van een referentiecorpus dat niet geheel aan de gestelde kwaliteitseisen voldoet. Om de betrouwbaarheid te vergroten zou daarom grotere en meer valide referentiecorpora gebruikt moeten worden, en zou in meer en gevarieerdere contexten vergelijkbaar kwantitatief onderzoek plaats moeten vinden. De laatste aanbeveling geldt ook

---

70 Zie http://corpus.byu.edu/coca/
71 Zie Boulton 2010.
voor het ontwerponderzoek aangezien deze de generaliseerbaarheid van het onderzoek namelijk aanzienlijk zou kunnen vergroten.

De conclusies van het onderzoek ondersteunen ook de noodzaak voor grootschaliger en uitvoeriger onderzoek naar het verschil in gebruik van connectieven in het Nederlands en in het Engels. Men zou systematisch de gedragingen van de belangrijkste connectieven op individueel niveau en vanuit contrastief opzicht in kaart moeten brengen.\textsuperscript{72} Vervolgens kan er zeer specifiek remediërend lesmateriaal ontworpen worden. Onderhavige onderzoek biedt de corpus tools, het analytisch model en een activiteitentypologie voor docenten en onderzoekers om daarmee op bescheiden schaal aan de slag te gaan.

Wat betreft het gebruik van corpora is een eerste experimentele stap gezet om leerlingen op een actieve en zelfstandige wijze aan hun taalontwikkeling te laten werken. Vooral waar het redactie en revisie van geschreven teksten betreft en het ondersteunen van tweetalige vaklessen, hebben corpora veel te bieden. Echter, het zelfsturend en -ontdekkend (taal)leren, en alle noviteiten en onzekerheden die daar bij horen, kent een Nederland (nog) relatief weinig traditie. De experimenteerder groep geeft, zoals was te verwachten, dan ook duidelijk aan het geheel zelfstandig selecteren van relevante gegevens uit grote hoeveelheden informatie, intimiderend te vinden, zeker als daar systematische werkhou ding aan ten grondslag moet liggen. Deze beperking zou in een toekomstige interventie deels kunnen worden ondervangen door het aantal te onderzoeken connectieven verder te beperken en nog meer op diepgang in te zetten. Bovendien moeten de vormgeving en gebruiksvriendelijkheid van corpora en corpustools meer toegespitst worden op de specifieke behoeftes van de doelgroep.

Tot slot, de constatering dat scholieren het lastig vinden om op gedisciplineerde wijze om te gaan met groeiende hoeveelheden (digitale) informatie, is inherent aan de \textit{multitaskende Einsteinengeneratie}. Het is daarom zaak dat het onderwijsveld zich inzet om hen een effectieve gereedschappen aanreikt teneinde ze zo snel mogelijk op productieve wijze grote hoeveelheden informatie te kunnen laten verwerken.\textsuperscript{73}

\textsuperscript{72} Zie als voorbeeld het onderzoek van Springer 2012.
\textsuperscript{73} Voor het belang van digitale informatie-ontsluiting, ook voor de geesteswetenschappen, in de toekomst zie de inaugurele rede van Prof. dr. Rens Bod
\url{http://www.english.uva.nl/news/professorappointments.cfm/3476083D-5913-4663-9E9F02DF06446A5A}
Anderson & Corbett 2009

Aijmer 2009

Andrewes 2011

Aston 2001

Aston et al. 2004

Baker et al 2009

Baker 1992

Ball 1986

Bennet 2010

Berzlánovich, Egg & Redeker 2009

Bikelienè 2008

Boulton 2010

Carrió 2006

Celce-Murcia, Larsen-Freeman & Williams 1999

Chalker 1996

Crewe 1990

Dale, Es & Tanner 2011

McEnery, Xiao & Tono 2006

Evers-Vermeul 2005

Gavioli & Aston 2011

Granger 2002

Granger, Hung & Petch-Tyson 2002

Granger & Tyson 1996

**Gray & Sutherland 2010**

**De Haan & Van Esch 2007**

**Halliday & Hasan 1976**

**Hatch 1992**

**Horn 1972**

**Horváth 2001**

**Hunston 2002**

**Kettemann & Marko 2002**

**Kosem 2008**

**Lackman 2010**

**Lew 2009**

Lieber 1980

Mahlberg 2006

Müller 2005

Marco 2010

Marks 2010

Mukherjee 2002

Mukherjee 2006

O’Keeffe, McCarthy & Carter 2007

Pander Maat & Sanders 2006 (a)

Pander Maat & Sanders 2006 (b)

**Partington 2003**

**Quirk & Greenbaum 1973**

**Ranalli 2003**

**Reppen 2010**

**Sanders 1994**

**Scarcella 1984**

**Schiffrin 1988**

**Scott & Tribble 2006**

**Shaw 2010**

**Siepmann 2005**

**Smith & Frawley 1983**

**Springer 2012**


**Stubbs 1996**


**Tanskanen 2006**


**Tavecchio 2010**


**Teubert & Čermáková 2007**


**Wong 1993**

Wong, T.Y. (1993) *Textual Cohesion and Reading Comprehension* [thesis MPhil in Education], Hong Kong.

**Zamel 1983**

Appendix 1: CONNECTIVES CHART

The chart contains connectives which have been categorised and subcategorised according to their logical function: (1) additive [and]; (2) adversative [but]; (3) causal [so]; (4) temporal [then]. The first three or four letters of each heading of the subcategories have been capitalised so that they can serve as easy-reference acronyms in checking or revising written work.

(1) ADDITIVE

(1A) **ADDitive connectives** add the following sentence to the previous sentence(s)

**Examples:** also; and; moreover; furthermore; another; besides; too; in addition; again; similarly; aside from; likewise; as well as.

(1B) **ENUmerative connectives** list elements of a catalogue.

**Examples:** at first; first of all; firstly; secondly; next; finally; last; (the) last; then; since then

(1C) **SUMmative connectives** introduce a sentence which summarises the previous one.

**Examples:** in conclusion; in summary; in short; to sum up; in one word; all things considered; it is concluded; in brief; my conclusion;

(1D) **APPositive connectives** introduce a reformulative sentence of the previous sentence(s).

**Examples:** that is; this means; in other words; that is to say; namely; I mean.

(1E) **EXAmples connectives** introduce a sentence which is an example of the previous sentence.

**Examples:** for example; for instance; especially; particularly; in particular.
(1F) **MANner connectives** indicate that the previous sentence describes the manner in which the contents of the following sentence is conducted.

   Examples: *in this way; like this.*

(1G) **TRAnsitive connectives** mark a change of subject.

   Examples: *now; by the way; incidentally.*

(1H) **REFerential connectives** indicate that the following sentence focuses on a particular point.

   Examples: *as for; as to; concerning; from the point of view; in terms of; speaking of; with regards to; regarding.*

(2) **ADVERSAIVE**

2A) **CORROborative connectives** indicate the writer’s conviction that the content of the following sentence is true.

   Examples: *of course; in fact; indeed; naturally; clearly; surely; in effect; as a matter of fact.*

2B) **CONCessive connectives** indicate that the following sentence is contrary to expectation.

   Examples: *however; but; albeit; yet; nevertheless; in spite of; after all; at the same time; in any case; (al)though; even so; anyway; anyhow.*

2C) **CONTrastive connectives** introduce a sentence which contrasts with the previous sentence.

   Examples: *on the other hand; some and other; in comparison; by contrast.*

2D) **CORrective connectives** introduce a sentence which corrects the previous sentence(s).

   Examples: *on/to the contrary; rather; instead; otherwise; far from that/it; or.*

3 **CAUSAL**
3A) **Causal connectives** indicate that the sentences in between them are cause and effect.

   Examples: therefore; thus; so; this/that/is is because; consequently; as a result; that/this is why; for these reasons; because of this; for; the reason; that it is a reason why; it was partly due to.

3B) **Inferential connectives** indicate that the following sentence can be inferred from the previous sentence.

   Examples: then; so.

4 **TEMPORAL**

4A) **Temporal connectives** indicate that the sentences between them are connected in time.

   Examples: (until) then; finally; later; this time; next; now; at last; (at) first; at that time; in the end; after that; from then on; from now; before this; prior to; since then; in the age; after; until now; at this time.

4B) **Local connectives** indicate that the sentences between them are connected in place.

   Example: here; in this case; from here on; on this point; up to this point.
Appendix 2: Profile of learner corpus

1) Type of corpus: Learner Corpus (NNS) or Reference Corpus (NS)
   LC (NNS)

2) Date of corpus production.
   11.11.2009

3) Input/source, with reference (e.g. students or internet).
   Students, 4V1

4) Total number of words of corpus (excluding punctuation marks).
   14.866

5) Type of discourse text
   Analytical literary essays

6) Number of participants and discourse texts in corpus.
   25/25

7) Average length of discourse texts in words.
   500 words

8) Average age of participants.
   15/16

9) Average language level of participants.
   Advanced/adolescents

10) Native-language of participants.
    Dutch (+ two native speakers of English)

11) Participant’s previous experience with language point (e.g. no, basic or advanced knowledge of connectives)
    Experience with cohesion and essay writing.

12) Additional information.
Appendix 3: Pre and post-test (of design research)

PRE-TEST: ACTION RESEARCH ON CONNECTIVITY

Name: ______________________      Group: ______                  Date: __________

Total score (out of 29 pts): _____       Time needed in minutes: ____

This short test is a pre-test and will be used for research on the use of connectives in formal texts written by adolescent students in the Netherlands i.e. YOU! It is a part of a research project in which you will participate as an experiment or control student. Please read the instructions below carefully and do the exercises to the best of your ability. *Good luck!*

**Section I: Semantics**

A) Identify and write down the connectives in each of the following sentences.  
(CON/002/3)

1) In terms of his natural skills, he had an excellent chance of winning the grand prize. Unfortunately, he did not have enough money to enter the tournament. Consequently, he never had the chance to prove himself and others that he was a serious competitor.
   Connectives: __________________________________________________________

2) It seems clear that we have no alternative except to go back and run this experiment again. Incidentally, one way to avoid the kind of problem that is forcing us to do this is to make sure that we follow basic procedures for assigning subjects to groups and that we follow the experimental design.
   Connectives: __________________________________________________________
B Link the two sentences basing your answer on the connective given. You can only use each connective once. Write the letter behind the sentences. (CON/010/9)

1. Chicago is one of America’s greatest city. However, ___
2. Chicago is one of America’s greatest city. In addition, ___
3. Chicago is one of America’s greatest city. Therefore, ___
4. Our new maths teacher is very professional. However, ___
5. Our new maths teacher is very professional. In addition, ___
6. Our new maths teacher is very professional. Therefore, ___
7. My uncle and aunt decided to leave for France. However, ___
8. My uncle and aunt decided to leave for France. In addition, ___
9. My uncle and aunt decided to leave for France. Therefore, ___

A) they didn’t particularly like the people there.
B) accommodation is terribly expensive.
C) he decided to make a detour.
D) it has excellent public transport.
E) our marks have gone up significantly this year.
F) he's quite humourless.
G) they lost contact with their relatives in the United States.
H) they sold their Dutch flat and everything of value.
I) people say she’s very good looking.

C Can you insert the appropriate connective in the following sentences? The category of the connective is given. (CON/003/5)

Our state’s correctional system is plagued with problems. (1) __________, (example) high officials increase their personal wealth by awarding building and catering contracts
to disreputable companies in return for bribes. (2) ___________, (addition) promotions within the system are made on the basis of politics, not merit. (3) ___________, (result) the system is filled with people at the top who know little about what they are doing. (4) ___________, (addition) careless security measures, allowing trusted inmates to control certain operations of the institution, are part of the growing problem. But one increasing tendency in particular is doing harm to the system’s image and efficiency. This is the tendency of officials who are charged with important tasks and who make faulty decisions to cover up their mistakes. (5) ___________, (conclusion) one would think that amid all the strife some effort would be made to rectify these problems, but a seemingly dogged determination to resist change overshadows the system.

Section II: Syntax

D The following (pairs of) sentences all contain connectives. Some of them occupy the wrong position in the sentence or lack appropriate punctuation. Can you identify these and correct their usage? Write down the full sentence and do not change the connective. (CON/017/3)

1) The 10:15 IC train to London broke down and also the next one.

______________________________________________________________________
______________________________________________________________________

2) Dad wants to spend the summer in northern Italy. On the other hand, my mother insists we drive down all the way to Sicily.

______________________________________________________________________
______________________________________________________________________

3) Throughout the book it does not become clear what this “prophecy” is about, however this creates a conflict in Elspeth whether to believe it and try to fulfil this prophecy or whether to keep her life going.

78
E Replace each of the following pieces of text with a new sentence or sentences as similar as possible in meaning to the original but using the connectives given in brackets. Make any other changes that are needed. (CON/005/5)

1) While I was reaching for the glove compartment she was getting out of the car. (meanwhile)

2) Though Tibetan was to be taught during the three years of primary school, all secondary education was in Chinese by Chinese teachers. (however)

3) Whereas Ireland worked as a team with clever passing among the forwards, England relied on individual efforts. (by contrast)

4) Whereas unemployment seems to have contributed to increasing crime rates, this has not generally resulted in more people being sent to prison for longer periods of time. (however)
5) Much as he had enjoyed the lively company of his friends in Edinburgh, his work, he felt, demanded solitude. (all the same)

Section III: Style/register

F The following sentences all contain one or more connectives. However, some of them are too informal for the context in which they are used. Locate the inappropriate connectives and replace them by more appropriate ones. (CON/018/4)

1) Mayer says a clergyman could argue that this is an economic hardship. Besides, the professor adds, "The government isn't allowed to condition privileges on the giving up of your constitutional rights, at least not absolutely."

2) Indeed, one of the tragedies of life is that so many deep ideas aren't true. At any rate, I am claiming that two of Plato's deep ideas about the nature of love are at work in our psychic depths and thus have a large and ongoing (if hidden) influence on our love lives.

3) Perplexed longshoremen were summoned to the levee to help move some enormously heavy crates marked "Tamaroa marble": material for an upcoming art exhibition, they were told. Actually, the crates contained two howitzers, two siege cannons, 500 muskets, and a large supply of ammunition, all recently confiscated by Confederate authorities from the U.S. arsenal down in Baton Rouge.
4) The novel is read by almost all British students at secondary school. So it must have some specific qualities which other novels lack.

POST-TEST: ACTION RESEARCH ON CONNECTIVITY

Name: ______________________      Group: ______                  Date: __________

Total score (out of 29 pts): _____     Time needed in minutes: ____

This short test is the second (post-) test on the use of connectives. It is similar in structure to the pre-test you took earlier. The results will be used to see if and to what extent you have improved your connectivity skills. It is a part of a research project in which you will participate as an experiment or control student. Please read the instructions below carefully and do the exercises to the best of your ability. Good luck!

Section I: Semantics

A Identify and write down the connectives in each of the following sentences. (CON/002/3)

1) As it had snowed heavily all day long for the past week they were forced to stay in the cabin. Still, they refused to give up hope that the sun would come out sometime in the next three days and that they would at least get in some skiing before their vacation was over.
Connectives: ____________________________________________________________

2) I had expected that she would at least speak up and offer some support in my defence. Instead, she remained silent during the attack that was coming from the other members of the committee.
Connectives: ____________________________________________________________
B  Link the two sentences basing your answer on the connective given. You can only use each connective once. Write the letter behind the sentences. (CON/010/9)

1). It rained cats and dogs last autumn. Consequently, _________
2). It rained cats and dogs last autumn. Nevertheless, _________
3). It rained cats and dogs last autumn. For example, _________
4). The new college attracted more students than expected. Consequently, _________
5). The new college attracted more students than expected. Nevertheless, _________
6). The new college attracted more students than expected. For example, _________
7). My latest contribution to the blog was too long, I felt. Consequently, _________
8). My latest contribution to the blog was too long, I felt. Nevertheless, _________
9) My latest contribution to the blog was too long, I felt. For example, _________

A) on one October day more than 2 inches of water came down.
B) lecture halls were divided into smaller rooms.
C) the fire brigade was called out to help umpteen times.
D) people sent emails complaining about the amount of time it took to read.
E) the kids continued to play outside almost all days.
F) I noticed that not all students had access to lockers.
G) my neighbour’s father is still out of a job.
H) the webmaster said that he wanted new texts to be published.
I) those paragraphs that were uploaded didn’t make sense.

C  Can you insert the appropriate connective in the following sentences? The category of the connective is given. (CON/004/5)

Genetic screening in business, or testing the genes of employees to see if they are susceptible to workplace-related diseases, may present problems for the tested. (1) __________ (example), the genetic screening tests and technology in general are in their
infancy stages. (2) _________ (result), many physicians and health professionals doubt their reliability. (3) _________ (addition), once genetic information is recorded on employees, it cannot always be kept secret. Even though employers are assured that their medical files are confidential, clerical staff have access to them. (4) _________ (example), if they are entered into a computer data base, they are available to anyone with access. (5) _________ (result), some argue that such screening procedures are violations of personal rights.

Section II: Syntax

D The following (pairs of) sentences all contain connectives. Some of them occupy the wrong position in the sentence or lack appropriate punctuation. Can you identify these and correct their usage? Write down the full sentence and do not change the connective. (CON/017/3)

1) He can be very polite and friendly to people, which can be very confusing. This is very sneaky of Alex, because some people actually trust him.

2) Some of Frodo’s friends are skilled fighters and they win most of the time. One of his friends is however killed by orcs.

3) There also is a conflict between Jerry and Charlie and their father, Allie.
Replace each of the following pieces of text with a new sentence or sentences as similar as possible in meaning to the original but using the connectives given in brackets. Make any other changes that are needed. (CON/005/5)

1) While some UN sources expressed hope that a deal would be struck, the Americans were less optimistic. (on the other hand)

2) Although they knew it was wrong to steal a car, most of those interviewed, all under 17, did not consider themselves criminals. (despite)

3) ‘I can’t do anything for several days,’ he said. ’There’s an important murder trial going on and I’m the main police witness. It won’t matter a lot, though.’ (still)

4) We had our money stolen on the last day. Otherwise we had a wonderful time there. (except that)

5) She is a very good English speaker. You would think it was her native language. (so)
The following sentences all contain connectives. However, some of them are either too informal for the context in which they are used. Locate the inappropriate connectives and replace them by more appropriate ones. (CON/018/4)

1) The other group is the ‘Eloi-race’ and they are afraid of the dark because then the Morlocks try to get them. Time, actually, is the subject of the novel.

2) Many students are well prepared with the field skills required for the position in which they were hired. The field skills of others may be comparable, but not as appropriate for their area of work. Also, the plant identification skills category may have been too broad.

3) For various local reasons - spiritual inspiration, promotion from within, exceptional talent, a family's increased wealth - a masker's function may change a little or entirely. Because of this, that mask and headdress may be changed or elaborated, or the spirit/performer may continue to prefer the original mask as it was.

4) Each well can hold one of the large DNA extraction beads. Then the plate is filled with enzymes that release a flash of light whenever a chemical called a nucleotide bonds with the DNA in a well, revealing a genetic code.

Pre-test key:

A  (3 pts)
1  [Unfortunately,]; [consequently] (2 pts); [and] is not relevant.
B (9 pts)
1 – B / C  
2 – D  
3 – C  
4 – F  
5 – I  
6 – E  
7 – A  
8 – H  
9 – G / H  

C (5 pts)
(1) *Example:* [For one thing,]; [Frequently,]; [For instance,]; [For example,]; [Often,]
(2) *Addition:* [Furthermore,]; [Moreover,]; [In addition,]; [What is more,]; [Also,] (informal but correct)
(3) *Result:* [As a result,]; [Consequently,]; [Because of this,] (informal but correct), [Therefore,]
(4) *Addition:* [In addition,]; [What is more,]; [Additionally,]; [Also,]
(5) *Conclusion:* [In short,]; [In conclusion,]; [To conclude,]; [In the end,] is wrong because is not a conclusion.; [So] is wrong because too informal; [Concluding,] frequency in COCA is very low.

D (3 pts)
1) [also]: The 10:15 IC train to London broke down and the next one also.  
2) [On the other hand]: Dad wants to spend the summer in northern Italy. My mother, on the other hand, insists we drive down all the way to Sicily.  
3) [however]: Throughout the book it does not become clear what this “prophecy” is about. However, this / This, however, creates a conflict in Elspeth whether to believe it and try to fulfil this prophecy or whether to keep her life going.

E (5 pts)
1) I was reaching for the glove compartment. Meanwhile, she was getting out of the car.
2) Tibetan was to be taught during the three years of primary school. However, all secondary education was in Chinese by Chinese teachers.
3) Ireland worked as a team with clever passing among the forwards. England, by contrast, relied on individual efforts. / By contrast, England ...
4) Unemployment seems to have contributed to increasing crime rates. However, this has not generally resulted in more people being sent to prison.
5) He had enjoyed the lively company of his friends in Edinburgh. All the same, his work, he felt, demanded solitude.

F (5 pts)
1) [Besides,] = [In addition,]; [Furthermore,] ([Also,] is informal and wrong.
2) [At any rate,] = [Whatever the case may be]; [Indeed] is not relevant here.
3) [Actually,] = [In reality,]; [In fact,]
4) [So] = [Therefore,]; [As a result,]; [Consequently,]

Post-test key:

A (3 pts)
1) [As,]; [Still,] ([and] is not relevant here)
2) [Instead]

B (9 pts)
1 – C
2 – E
3 – A
4 – B
5 – G
6 – F
7 – I / D
8 – H
9 – D / I

C (5 pts)
Example: First; First of all; For one thing; [Example] is wrong because no link with rest of text

2 Result: Consequently; Hence; Therefore; As a result

3 Addition: Second; Moreover; Further; What is more; Furthermore; In addition, ; Additionally,

4 Example: Indeed; Specifically

5 Result: As a result; Consequently; Hence; Thus; Therefore, ; Ultimately,

D (3 pts)
1) [because]: He can be very polite and friendly to people, which can be very confusing. This is very sneaky of Alex because some people actually trust him.

2) [however]: Some of Frodo’s friends are skilled fighters and they win most of the time. However, / One of his friends, however, is killed by orcs.

3) [also]: There is also a conflict between Jerry and Charlie and their father, Allie. / Also, there is ...

E (5 pts)
1) Some UN sources expressed hope that a deal would be struck. The Americans, on the other hand, were less optimistic.

2) Despite the fact that they knew OR Despite knowing it was wrong ... most of those interviewed ... did not consider themselves criminals.

3) .. ‘There’s an important murder trial going on and I’m the main police witness. Still, it won’t matter a lot.

4) We had a wonderful time there, except that we had our money stolen on the last day. [note: comma before [except] can be left out.]

5) She is a very good English speaker so you would think it was her native language. [Note: comma before [so] is acceptable.]

F (4 pts)
1) [actually] = [in fact]

2) [also] = [similarly]

3) [because of this] = [accordingly], [as a result of this], [consequently], [For this reason]

4) [Then] = [next], [Following this], [After this,]
Appendix 4: Student questionnaire about effectiveness of intervention on connectivity and corpus linguistics

The past few weeks we have been working with **connectives** and **corpora** (COCA, more specifically) to help you improve your language skills. We are now interested to find out how useful you think these materials, exercises and tools have been. Your responses to the statements and questions below will help us formulate conclusions and design new materials in the future.

**Note:** There is absolutely no need to 'please' your teacher with your answers so be sincere and constructive!

### Part I: Questionnaire

Please read through the statements below and rank your response on a scale of 1 to 5.

- **5** = 'I agree completely.'
- **2** = 'I hardly agree.'
- **4** = 'I agree for the greater part.'
- **1** = 'I absolutely do not agree.'
- **3** = 'I agree to some degree.'

#### Connectivity

<table>
<thead>
<tr>
<th>Statement</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) The various exercises on connectivity usage done in class have made me more aware of the importance of connectivity in written texts.</td>
<td></td>
</tr>
<tr>
<td>2) I found the exercises on the meaning of connectives (semantics) useful.</td>
<td></td>
</tr>
<tr>
<td>3) I found the exercises on the position of connectives and comma use (syntax) useful.</td>
<td></td>
</tr>
<tr>
<td>4) I found the exercises on the differences between <strong>formal and informal usage</strong> of connectives (style/register) useful.</td>
<td></td>
</tr>
</tbody>
</table>
5) I now feel more confident about the use of connectives.

Corpus tools and COCA

<table>
<thead>
<tr>
<th>Statement</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Corpora are <strong>useful tools</strong> to improve connectivity and general English language skills with.</td>
<td></td>
</tr>
<tr>
<td>2) Corpora are <strong>more useful</strong> in finding out about language than <strong>dictionaries</strong> are.</td>
<td></td>
</tr>
<tr>
<td>3) In the future, I will certainly <strong>use COCA</strong> (or any other relevant corpus) to help improve my language skills.</td>
<td></td>
</tr>
<tr>
<td>4) Corpora are <strong>very useful</strong> new tools that students can use to improve their language skills.</td>
<td></td>
</tr>
<tr>
<td>5) I now feel <strong>more confident</strong> in using corpora to help me improve my language skills.</td>
<td></td>
</tr>
</tbody>
</table>

Part II: Open questions

1) What would you consider the **greatest benefit** of corpora (such as COCA)?

2) What was **most interesting** to you about the use of corpora?
3) What aspect of corpora did you think was most difficult? And why?

4) What do you think is the greatest disadvantage to corpora?

5) Any further comments or suggestions?

Many thanks!
Using corpora in the classroom
A corpus approach to connectivity

Introduction
In this session you will be introduced to and work with a relatively new but extremely useful area in language learning and teaching: web-based corpora with which you can further improve your English language skills, vocabulary and grammar. This project focuses on one important language aspect: connectivity. You will find out how certain connectives tend to ‘behave’ in language i.e. what positions in a sentences they occupy, what relationships between sentences and clauses they denote, and whether they are mostly used in spoken or written language. In the first part your teacher will guide you through some theoretical and technical aspects of corpora and corpus tools; in part 2 you can work on your own doing exercises and using web corpora. The aim of the project is for you to find out how you can use corpora to improve your language skills independently.

Part I: Theory

What is a corpus?
A corpus (plural: corpora) is basically ‘a very large, principled collection of naturally occurring spoken or written texts in a certain language’. Corpora are composed using textbooks, fiction, magazines, academic papers, telephone conversation, amongst many other sources. A corpus is accessed and analysed using specific digital corpus tools and computer programs. The branch of linguistics which uses and creates corpora is called corpus linguistics.

Exercise 1: Some people say that Google is a corpus tool that can be used to analyse texts and language that you find on the web. What, according to you, are the main differences between Google and the corpora as defined above?
How can corpora help improve my English?
A growing number of linguists are primarily interested in seeing how language *is used* by people in real-life situations (rather than telling people how language *should be used*). For this they compile and utilise written or spoken corpora. But language learners like you can also benefit a lot from using corpora.

**Exercise 2:** Think of one way in which corpora might help you improve your language skills. Go back to the definition of corpus if you find this question difficult to answer.

What do corpora NOT do?
Corpora do not give ready-made answers. It is you, the user and learner, who will have to find the answer by using a clear research question (*What do I want to know about the behaviour of a certain word?*), locating the required corpus data (*Where and how can I find relevant information in a corpus?*) and actively creating patterns and connection (*How is this word commonly used by people?*).

**Exercise 3:** The following terms are often used in corpus linguistics. Use the internet and find clear definitions and examples. Make sure your definitions and examples are relevant to the context of corpus analysis (and not to any other contexts!).

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concordance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collocate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Part II: Practice**
What do I have to know to get started?

Using corpora may seem a little daunting because you have to know a few technical ‘tricks’ to get started and utilise corpora effectively. It’s a bit like knowing how to use a dictionary properly so you have to be patient and persevere. There are many corpora to be found on the internet, many of them providing specialised language and tools. We will be using a ‘general English’ corpus, the Corpus of Contemporary American English, commonly known as COCA, which was designed by professor Mark Davies from Brigham University in the USA. With over 400 million words (and growing every day) the COCA is the largest freely accessible corpus on the web. The COCA has excellent built-in tools (called a ‘concordancing program’ or simply ‘concordancer’) so there is no need for separate programmes. Please find a computer and access the COCA through www.americancorpus.org and follow the manual below.

How do I use COCA?

Access COCA and you will arrive at the introductory (blue) page. On the right you find some background information, and on the left there are five search fields, some of which we will be using. At this stage we are mostly interested in KWIC, a KEYWORD IN CONTEXT, which is the word we are going to research.

Exercise 4: In the COCA search form WORD(S) type in [however] and click on SEARCH. How often does [however] occur in COCA?

Exercise 5: Under the button CONTEXT click on HOWEVER (in blue). You will now see [however] in context. How many different sources have been used for the first 100 hits?

Exercise 6: In the list, click on number 1. What additional information about the KWIC is given compared to the previous screen? Why is this important?
Exercise 7: Return to KWIC entries by clicking the button at the bottom. As you may know, [however] has two completely different meanings. Read the first 10 lines and find two example sentences for each of the two meanings.

1:
2:
3:
4:

Exercise 8: On the basis of the example sentences please try to formulate what two meanings or functions [however] can have.

1:
2:

Exercise 9: Briefly describe any patterns you see in the position of commas in relation to the usage of (the two types of) [however].

You now have a basic working knowledge of corpora and COCA. In the remainder of the project you will use COCA independently to explore connectives and improve your usage of them. But always please ask your teacher in case of problems.

- Good luck!