

VU Research Portal

Advanced Learner Writing

Springer, P.E.

2012

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Springer, P. E. (2012). *Advanced Learner Writing: a corpus-based study of the discourse competence of Dutch writers of English in the light of the C1/C2 levels of the CEFR*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Chapter 7 Conclusion

7.1 Introduction

This final chapter will first summarize and briefly discuss the main research findings of the current study in relation to the research aims as they were outlined in the Introduction (7.2). To encompass the whole width of the study, the section will first start with the study's central research findings, the results of the corpus analysis (7.2.1). Next, these corpus research findings are related to the small-scale study of academic writing textbooks (7.2.2). The section ends by relating these findings to the CEFR, which was the starting point for the analysis of advanced learner writing (7.2.3). The chapter ends with formulating some methodological limitations of the current study and the interpretations based on it (7.3), as well as making suggestions for further research (7.4) and outlining some practical implications of the current research (7.5).

7.2 Main Findings

This study started with determining what the salient features of writing at the C1/C2 levels of the CEFR are. It was determined that what is asked of proficient learners is a well-developed discourse competence. Discourse competence was defined as 'the knowledge of and the skill to use the lexico-syntactic means available in a language to produce a smoothly flowing, cohesive and coherent text in a given genre by logically structuring sentences and the use of appropriate cohesive devices'. It was further determined that in order to chart the discourse competence of the learners lexico-grammatical devices belonging to the following four categories be annotated in the learner corpus, attention getting /focusing devices, backgrounding /clause combining devices, cohesive devices, and stance.

An important finding of this study is that the Dutch advanced learner writers, with regard to these categories, are in many respects very close to professional native speaker writers. For example, with regard to linking adverbials there is no significant difference between the learners and the NS writers in terms of relative frequency and positional features. This sets them apart from many other L2 writers described in the literature who seem to overuse connectors in general and initial position specifically to the detriment of other possible positions in the sentence (cf. Granger 1996; Gilquin et al. 2007a). But also in other areas it was often necessary to do a thorough quantitative analysis, often paired with a qualitative analysis, before any differences between the learners and the professional writers came to light.

7.2.1 Main corpus research findings

The corpus analysis uncovered very different areas of overuse and underuse, some significant but quite subtle (e.g. overall learners underuse *-ed* clauses, but they overuse *-ed* clauses in initial position), others significant and very clear (e.g. the overuse of *because* clauses and other devices for indicating causal meaning relations). Chapter 4 presents all the quantitative findings of the corpus analysis and summarizes them in Table 101 on page 158. Some trends were identified with regard to semantic categories, position and syntactic complexity. Chapter 5 subsequently provided a discussion of selected results, as well as a qualitative analysis of some of the findings. The outcome of that analysis provided a basis for a systematic grouping of instances of overuse and underuse in four dimensions.

One of the clearest outcomes of the quantitative analysis is that learners overuse a number of lexico-grammatical devices that express a causal relationship, namely adverbial clauses of reason and linking adverbials of result/inference, such as *therefore* and *because of X*. Additionally, lexical formulae with a similar function, such as *this means that* and NCDUS

(see 5.4.4 above) starting with *because of* are overused by the learners. Another clear overuse concerns indicators of sequence, such as linking adverbials of addition, adverbial clauses of time (both finite and non-finite) and the overuse of another as a determiner in retrospective labels.

With regard to position, it is clear that in most cases learners have a preference for initial position when the choice presents itself where to put an additional constituent. This is exemplified by the clear overuse of non-restrictive relative clauses and non-restrictive nominal appositions in the theme zone and the relative overuse of initial position of non-finite clauses (meaning non-finite clauses are less underused in initial position than in medial or final position). Medial position is underused and this dispreference of interrupting the flow of discourse using additional, parenthetical comments is also reflected in the underuse of parenthetical clauses.

There are also some important differences between the learners and expert writers with regard to sentence complexity. Sentences in VUNSPRAC tend to be both longer and more complex than those in VULCAN. Moving from VULCAN_1 to VULCAN_3 to VUNSPRAC_ling to VUNSPRAC_lit, the mean clause length for all the annotated clause types increases, and with it the mean sentence length. Conversely, the same route sees a decrease in the number of simple clauses, i.e. clauses that contain none of the lexico-grammatical backgrounding/clause combining devices, such as adverbial clauses and appositions, and an increase of sentences that contain three or more of these backgrounding devices. This trend of VULCAN_1 < VULCAN_3 < VUNSPRAC_ling < VUNSPRAC_lit, or VULCAN_1 > VULCAN_3 > VUNSPRAC_ling > VUNSPRAC_lit is visible in many of the lexico-grammatical devices under investigation in terms of observed frequencies. This schematic also implies that there are differences between VUNSPRAC_ling and VUNSPRAC_lit. This is indeed the case. For example, VUNSPRAC_lit uses sentence initial *and* and sentence initial *but* more than twice as often as VUNSPRAC_ling. However, it was established that although it is important to note that VUNSPRAC_ling and VUNSPRAC_lit differ significantly in some respects (enough to always bear in mind that these are two different sub-genres of academic writing), these differences did not constitute a fundamental problem in the quantitative analysis of underuse or overuse by the learners.

Chapter 5 postulated four dimensions which systematically grouped the instances of overuse and underuse in terms of discourse competence. They are

- 1) *concision*: from an overuse of full forms, which spell out explicitly what the writer means, to a balance between full and compact forms, without losing clarity;
- 2) *authorial confidence and involvement*: from generality to interactivity and specificity, by interrupting the discourse utilizing all possible positions in the clause;
- 3) *sophistication*: from less to more sophisticated, both cognitively, i.e. writers at the sophisticated end of this dimension use a greater array of semantic relations –and therefore achieve a greater degree of semantic precision– than learners who are at the ‘simple’ end of the same dimension, and textually, i.e. certain text-types offer more opportunities for using specific lexico-grammatical devices;
- 4) *degree of formality*: from speech-like to appropriate in a formal written context.

Dimension 1, concision, is typified by the overuse of finite adverbial clauses and the underuse of non-finite adverbial clauses as well as the overuse of non-restrictive relative clauses and underuse of non-restrictive nominal appositions and ascriptive NP NCDUS. Dimension 2, authorial confidence, is typified by the relative overuse of initial position of non-finite clauses and the underuse of medial position, parentheticals, reporting clauses, *as* clauses, elaboration, and ascriptive NP NCDUS. Dimension 3, sophistication, is typified by the underuse of concession (both finite adverbial clauses and linking adverbials), underuse of reporting clauses and a limited range of stance adverbials. The overuse associated with this dimension is apparent in the overuse of sequencing devices such as clauses of time and linking

adverbials of addition, the overuse of devices indicating cause or reason meaning relations, the overuse of focusing devices, the overuse of finite adverbial clauses of condition and purpose, and the overuse of vague nouns in retrospective labels. Dimension 4, degree of formality, finally, is typified by the underuse of apposition markers *e.g.* and *i.e.* and the overuse of the apposition marker *like* as well as the overuse of sentence initial *and* and *but*, (*even*) *though*, *while* as a subordinator of time (not concession), overuse of *wh*-clefts, indicators of causal meaning relations, and informal phrases such as *a lot of* and *lots of*.

7.2.2 Relating research findings to writing material

These four dimensions were subsequently used as a basis for the analysis of current pedagogical material for academic writing. Since some of these findings that identify the differences between advanced learner writing are so subtle that many teachers are not aware of them on a conscious level, it cannot be expected that learners simply pick these up from wide exposure to academic writing. In fact, it has been argued that learners will generally not pick up even more obvious characteristics of academic writing by mere exposure (Cobb 2003; Hinkel 2003, 2004). Explicit instruction is needed therefore. Many studies have confirmed the effectiveness of explicit instruction in SLA (Norris & Ortega 2000). This is where teaching material, more specifically the academic writing textbook, comes in.

In general, writing textbooks have been judged quite harshly in the literature (e.g. Kaszubski 1998) and Gilquin et al. 2007a). Harwood (2005) provides an overview of what he calls the anti-textbook view. The main criticism is that very little, if any, of applied linguistics findings are included in the textbooks. Gilquin et al. (2007a) express it thus:

While materials designed to help students improve their academic writing are legion (e.g. Bailey, 2006; Hamp Lyons & Heasley, 2006) few are corpus-informed, relying instead on material writers' perceptions of what good academic writing is or should be. Because of this lack of empirical support, many of these tools provide misleading information and unsound advice, as comparisons of published EAP materials and actual usage reveal (cf. Paltridge, 2002).

The next phase of the investigation therefore focused on the degree to which recent textbooks on academic writing treat the lexico-grammatical discourse devices, the underuse and overuse of which are instrumental in the four dimension of discourse competence, and if so, how they are treated in these textbooks.

The four dimensions of discourse competence are, of course, a construct of this study and thus cannot be expected to be present as such in current EAP teaching material, but by making an inventory of which lexico-grammatical devices are treated in the textbook it is possible to link these to the different dimensions, thereby describing which are present and how.

Unsurprisingly, the most 'common-sensical' of the four dimensions, Dimension 4 (degree of formality), was the only one that was actually identifiable as such in the textbooks under analysis and all textbook devoted some space to it. I use the term 'common sense' here, as the formal character of academic writing is its most widely reported feature, as well as learners' (both L2 and L1 novice writers) difficulty with acquiring this formal register. Of course, determining what the precise discrepancies between learner writing and expert writing in this area are is helped by corpus research. Most of the textbooks did not seem to use corpus research evidence for their claims, and if they did, they were mainly informed by L1 corpus research. Gilquin et al. (2007a) came to the same conclusion when investigating EAP writing material. So although all the textbooks highlighted the difference between spoken and written language, or formal and informal and mentioned relevant features of both, not all of them

mentioned the lexico-grammatical features that the current study found to be included in Dimension 4.

Features of the other three dimensions were much less likely to be treated in the academic writing textbooks under investigation. With regard to Dimension 1, concision, lexico-grammatical devices that might enhance the learners' syntactic efficiency, such as verbless clauses, appositions and NCDUS are ignored by most textbooks. Many of the textbooks do not discuss non-finite clauses, unless to warn learners against misuse in the form of 'dangling participles'. With regard to Dimension 2, authorial confidence and involvement, learners are not presented with the options for interrupting, specifying or elaborating the discourse by strategically placing subordinate structures in medial or final position, thereby asserting their authority and involvement. In other words, most textbooks do not explore the discourse potential of varying the position of lexico-syntactic features in the sentence. Nor are parenthetical clauses (also an instrument to interrupt the discourse, or add additional information to it) discussed by any of the textbooks. Conversely, reporting clauses and the closely related *as*-clauses are discussed in most textbooks, albeit in varying degrees of depth and accuracy. Linking adverbials of elaboration are not treated extensively, despite the fact that linking adverbials are presented in most textbooks, albeit without much explanation or context. With regard to Dimension 3, sophistication, the majority of textbooks did not discuss clause combining, let alone the different semantic categories of clauses that are underused or overused by the learners. Textbooks that do discuss semantic categories do not seem to introduce any specific ordering with regard to frequency of occurrence in actual academic writing or pointers with regard to overuse or underuse by learners. Only one textbook discussed retrospective labels, but makes no distinction between technical and vague head nouns. The overuse of linking adverbials of addition, another feature of Dimension 3, seems to be potentially reinforced by the fact that linking adverbials of addition are usually presented first of all linking adverbials, usually without context or ordering according to frequency, thereby receiving unintentional prominence.

Obviously, there are many important issues that are discussed in these textbooks and much useful advice given to the advanced learner. However, despite the present existence and accessibility of corpora, relatively little seems to have found its way into these types of textbooks. Learner corpus research, particularly, might make an important contribution to further enhance academic writing textbooks. Most, if not all, of the findings of the corpus research of the current study could be used to teach advanced Dutch writers of English to become even better writers (see also 7.5 below). It might not be financially feasible, or even desirable from a reader's viewpoint, to revise current pedagogical in such a way as to incorporate all the different idiosyncrasies of different learner languages from different backgrounds. However, an important step would be to use native academic English corpora to describe accurately what the core of academic English looks like and, next, to use learner corpus research to describe how that differs from most learners' interlanguage (cf. Gilquin et al. 2007b). For language-specific instruction, based on the learners' L1, extra web-based material could be made available to supplement the written material that represents the core course.

Finally, Leech (1998: xx) and Granger (2009: 22-23) contend that not every learner corpus research finding should necessarily feed into the development of teaching materials. Granger (2009: 22) states that 'whether features are selected for pedagogical action or ignored depends on a variety of features, including learner needs, teaching objectives and teachability.' However, it seems plausible that for any advanced learner trying to achieve near-native proficiency there are not many, if any, discrepancies between learner language and expert performance that can safely be ignored. It follows that the objective of EAP teaching material should be to address the needs of learners trying to achieve the highest possible proficiency in academic writing. In other words, there is not much that these learners do not have to learn, or that their teachers can choose to ignore, nor indeed the textbooks on

academic writing. Moreover, the starting point of this study was the determination of competences that very advanced learners should achieve, based on the description of the ‘can do’ statements in the CEFR. These descriptors are currently widely used as an ordering principle for newly developed teaching material.

7.2.3 Relating research findings to the CEFR

The CEFR states (Council of Europe 2001: 131) that

in order to participate with full effectiveness in communicative events, learners must have learnt or acquired:

- the necessary competences ...
- the ability to put these competences into action...
- the ability to employ the strategies necessary to bring the competences into action.

Of course, the focus of this study has been effective academic writing and in Chapter 2 of the current study two important conclusions were drawn with regard to the necessary competences: to write at the highest levels a learner has to have a well-developed discourse competence and, within the concept of communicative competence, linguistic competence is central and essentially feeds into discourse competence. The CEFR also states that ‘[t]he development of the learner’s linguistic competences is a central, indispensable aspect of language learning’ (Council of Europe 2001: 149) and ‘[g]rammatical competence, the ability to organise sentences to convey meaning, is clearly central to communicative competence’ (2001: 151).

Linguistic competence, of which grammatical competence is a component, is central, as a learner needs to have acquired a broad lexico-grammatical repertoire with a great semantic depth in order to be able to ‘write clear, smoothly flowing, complex texts in an appropriate and effective style and a logical structure which helps the reader to find significant points’ (Council of Europe 2001: 61).

It is possible to map the four dimensions of discourse competence onto this descriptor of overall written production at the C2 level, as it were, as can be seen in Table 119:

Table 119: Mapping the four dimensions of discourse competence onto the CEFR

<i>Dimension</i>	<i>name</i>	<i>CEFR C2 descriptor for overall written production in parts</i>
3	<i>sophistication</i>	‘Can write clear, smoothly flowing complex texts’
4	<i>degree of formality</i>	‘in an appropriate’
1	<i>concision</i>	‘and effective style’
2	<i>authorial confidence</i>	‘and a logical structure which helps the reader find significant points.’

Arguably, Dimension 3, *sophistication*, is related to the ability to ‘write clear, smoothly flowing complex texts’. Dimension 4, *degree of formality*, deals with ‘an appropriate style’, while Dimension 1, *concision*, could be seen as the dimension that deals with ‘effective style’. The ‘logical structure which helps the reader find significant points’ falls under Dimension 2, which has as one of its defining characteristics the use of position to order information in the sentence.

The corpus analysis of advanced learner language and the subsequent analysis of current textbooks on academic writing has shown that in order to facilitate learners to achieve the highest levels of proficiency in writing, as described by the CEFR, they might consider integrating some of the corpus findings from the current study as well as those of other learner

corpus studies in their teaching materials. Developers of pedagogical material should then consider which lexico-grammatical structures should be presented to the learners and ultimately be mastered by them, and are indeed encouraged by the CEFR to do so (2001: 153), in order to broaden and deepen their linguistic repertoire in such a way that they are skilled enough linguistically to be highly effective textually.

7.3 Limitations

During the course of this study, which started with an analysis of the descriptors of language competence in the Common European Framework of Reference and ended with an analysis of academic writing textbooks –and a corpus study at its heart– a wide range of decisions had to be made. These decisions have inevitably led to a focus on certain aspects while disregarding others. This section will briefly describe the limitations of the central corpus study, starting with the methodological limitations and ending with the limitations in relation to the interpretation of the results.

7.3.1 Methodological limitations

Most of the methodological limitations are related to choices that were made as a result, directly or indirectly, of the sheer amount of time it takes to manually compile, annotate and analyse a corpus of almost 300,000 words. Time constraints did not permit the corpora used in the current study to exceed their current size or more lexico-grammatical devices to be analysed, nor sometimes at the ideal level of detail.

Ideally, both VULCAN and VUNSPRAC would be larger than they are now. Although they are representative samples of the language varieties under investigation and larger than most corpora used in studies cited in this thesis, they turned out to be too small, nonetheless, for the accurate analysis of rarely occurring lexico-grammatical discourse devices, such as object postponement, subject–dependent inversion, or verbless clauses. These relatively rare devices require a much larger corpus. Furthermore, some of the more frequently occurring lexico-grammatical devices, such as finite adverbial clauses, could not be analysed statistically in more detail, for instance by statistically testing a three-way interaction between corpus, semantic category and position, because as subcategorization increased, the number of occurrences dropped very quickly, presenting a similar problem to the rare devices with regard to an accurate analysis.

Looking at the learners' discourse competence, as opposed to just analysing learners' use of the passive, for instance, meant that the net had to be cast out wide, as it were. In other words, the study could not limit itself to the analysis of just one lexico-grammatical device. For each discourse feature type, attention getting / focusing devices, backgrounding / clause combining devices, and cohesive devices / lexical cohesion markers, choices had to be made, with regard to which lexico-grammatical devices to include and exclude in the analysis. Inevitably, due to time constraints, not all possibly relevant features were annotated. A notable example is the choice to exclude the annotation of the occurrence of embedded clauses. However, the lexico-grammatical features that were included, arguably, cover such a wide range that they make it possible to perform a relevant analysis of the discourse competence of Dutch advanced learners.

Finally, the great number of lexico-grammatical devices, and their subcategorization, included in the analysis meant that in the current study had to limit itself –for the most part– to quantitative analyses. Only selected results were analysed qualitatively in more detail. Of course, the analysis of the corpora under investigation would have been even more thorough had time and resources permitted to supplement the extensive quantitative analysis with an even more detailed qualitative one of all the items under investigation.

7.3.2 Limitations in relation to the interpretation of results

Apart from the methodological limitations there are also a number of limitations with regard to the interpretation of results that present themselves. These concern both the representativeness of the corpus and the statistical analysis.

The learner corpus research described in this thesis has as its object of study texts produced by advanced Dutch learners of English from one Dutch university. It is important to always bear in mind which language variety a specialized corpus represents and restrict claims based on research of this to that variety and not make claims beyond that. For instance, findings from VULCAN might tentatively be extended to all advanced Dutch learners of English, but not to all advanced learners of English. We have seen throughout the study that there are at least as many differences as there are similarities between different varieties of advanced learner language. The findings of this study do, however, provide a valuable basis for the comparison of advanced Dutch learner English with other varieties of advanced learner English.

The other limitation pertains to the statistical analysis of the observed frequencies of the lexico-grammatical devices under investigation. The quantitative analysis of the corpora always started with analysing VULCAN and VUNSPRAC as a whole and subsequently breaking the analysis down to their respective subcorpora. However, even when frequencies of subcorpora are analysed, one works with aggregated frequencies. The underlying assumption is that there is no great internal variation between the texts in the corpus, as they have been added using strict criteria for inclusion representing a particular language variety. Indeed, this is how the majority of corpus research is conducted. However, Flowerdew (2010), for instance, does not just look at the learner corpus as a whole, but also takes into account the differences between individual texts and concludes that the text-based frequencies give a more accurate and less favourable measure of learners' ability to use signalling nouns than the overall frequencies (2010: 47). Based on this difference he concludes that

[T]he different results of the aggregated and individual text-based frequencies may have methodological implications for corpus study in general, suggesting that aggregated statistics may mask internal variations in a corpus and even misrepresent the general tendency under the effect of outliers. Greater significance, it follows, should be attached to analysis on the individual text level within a corpus, especially when individual variations are assumed, as is often the case with learner corpora. (2010: 48)

Arguably, the inclusion of many texts, each by different writers, decreases the influence of outliers in this respect. I have also tried to minimise this type of bias by checking, in the case of surprising or extreme results, if there seemed to be an individual text responsible for a particular outcome (using WinGrep), not just in VULCAN, but also in VUNSPRAC, as, arguably, individual variation is not just a feature of learner writers' texts, but of any type of writing that allows for personal stylistic variation. Of course, ideally, preventing this type of bias should be done in a more objective way. Mulder (2008: 43-56) explains how Multilevel Modeling (MLM) might be used to test statistical assumptions regarding the comparability of individual texts within a corpus.

As in most other corpus studies, the quantitative analysis used in the current study does not take the internal variations in the corpora into account statistically. MLM might be a useful methodological addition to future corpus research to ensure even greater representativeness of statistical outcomes of quantitative analyses of relative frequencies.

7.4 Suggestions for further research

The outcomes of the current study as well as its limitations and, indeed, the corpora that were annotated for the purpose of this study provide a basis for further research. As the scope of this study is very wide, many interesting opportunities present themselves.

The first extension of this research that comes to mind is replicating it using texts written by learners with a different L1, e.g. Spanish (or Norwegian, or Czech) advanced learners of English. By comparing the results with those of the current study more insight might be gained into the role of transfer as well as common features of learner language. This in turn might lead to more accurate advice being given to all advanced learners in those areas they have in common and a more targeted approach for those phenomena that occur due to the L1 background of learners. Similarly, the current research could be replicated using a corpus of texts written (under similar conditions) by NS student writers, in order to establish which differences with the professional writers' texts are based on cognitive, or developmental differences, regardless of L1 background.

Another way of extending the research would involve enlarging both VULCAN and VUNSPRAC. Some lexico-grammatical devices, most notably some of the rarer focusing devices, could not be investigated thoroughly due to their rarity combined with the size of the corpus. Using larger corpora might make it possible to analyse the relative frequency of even these relatively rare lexico-grammatical devices, as well as the infrequently occurring semantic categories of common devices, with much more accuracy.

Investigating the overuse and underuse of lexico-grammatical devices is a quantitative exercise. The qualitative analysis may reveal further differences between NNS advanced writers and NS writers. Another suggestion for further research might therefore be a more thorough qualitative analysis of the occurrence of the lexico-grammatical devices under investigation. Chapter 5 presented a discussion and qualitative analysis of selected results, but of course other devices also merit further investigation. For example, it was shown that in terms of positioning and frequency the learners use linking adverbials almost exactly like the NS writers. It would be interesting to analyse whether the learners use them accurately in terms of their function and semantics.

The current study has focused on the difference in discourse competence between advanced NNS learners and expert NS writers. However, there have been indications during the course of this study that suggest that the 'foreign soundingness' or the less than optimal effectiveness of learner texts might also be partly attributed to other factors, such as the use of collocations (e.g. Altenberg 2001; Barnbrook 1996; Howarth 1998), technical academic vocabulary (e.g. Coxhead & Byrd 2007; Gilquin et al. 2007b), and use of hedging devices (e.g. McEnery & Kiffle 2002). Investigating one or more of these factors using the corpora that were compiled for the current research might shed some light on why texts that are error-free still might strike the reader as being written by a NNS writer. The use of collocations could be explored using the n-grams functionality in Wordsmith and analysing the relative frequencies in the corpora as well as their semantic functions. With regard to academic vocabulary, the keywords of all corpora could be compared to the academic wordlist (Coxhead 2000, 2004) and analysed.

7.5 Practical implications

In section 1.5 it was stated that the current research aims to inform writing theory, CLC research and EAP. Arguably, implications of the current study extend to all three of these disciplines.

As regards writing theory, the main areas in which this thesis has attempted to make a contribution is with regard to the interpretation of the concept of discourse competence. So far studies that have used the term *discourse competence* (e.g. Alonso Belmonte & McCabe 2004; Celce-Murcia, M., & Olshtain, E. 2001; Ehrlich 1988; Hannay 2007a; Mauranen 1996),

do not provide a clear definition of the term. They either adopt, usually implicitly, Canale's rather broad definition, 'mastery of how to combine grammatical forms and meanings to achieve a unified spoken or written text in different genres' (1983: 9), assume that the reader knows what is meant by it, or provide a list of notions that all fall under the umbrella term *discourse competence*. A prime example of the latter is the description of discourse competence in the CEFR (Council of Europe 2001: 123). Building on the different descriptions of discourse competence found in the literature (mainly, it seems, Bachman (1990) and Canale (1983), but no reference is provided), this study has provided a new definition of discourse competence that is at the same time a more detailed update of Canale's (1983) definition and line with the CEFR's description. Furthermore, the notion of discourse competence was translated into specific discourse feature types and subsequently into a wide array of specific lexico-grammatical devices. By not looking at just one feature, such as thematic development (e.g. Mauranen 1996), or cohesive devices (e.g. Ehrlich 1988), but rather taking a whole range of features into account, a more complete picture emerges of how a theoretical concept such as discourse competence translates into real language, in this case English academic writing. Furthermore, by taking the omnipresent CEFR as its starting point and clarifying some of the concepts within it, this thesis may have helped clarify what the C1 and C2 levels of writing might entail. This is useful for language professionals who struggle to align their pedagogical material with the CEFR's descriptors. And, although this thesis has looked at the CEFR and at discourse competence primarily from a pedagogical angle, the analysis of the CEFR's highest-level descriptors for writing may provide useful insights for the field of language assessment research.

The contribution of this study to the field of CLC research is threefold. First, the current study has provided a model for investigating discourse competence in CLC and provided an annotation scheme. The corpus annotation was set up in such a way that it is readily applicable to any other learner corpus for the investigation of discourse competence. Many CLC studies focus on one specific lexico-grammatical element, whereas adopting the current methodology allows for a much broader picture. Of course, the annotation, as indeed the corpus compiled for this thesis, can also be used to investigate the occurrence of single items. Another contribution that this study has made is the uncovering of many instances of overuse and underuse in Dutch advanced learner English, which hitherto were unknown to the CLC research community. These results may benefit researchers working on other varieties of advanced learner language, as they can use the outcome of this study for their contrastive interlanguage analysis, comparing their L2 data with both L1 data and L2 data from the current study. Finally, CLC research aims to inform, among other things, research into second language acquisition and language pedagogy. By identifying –using learner corpus research– the four dimensions along which the progress of an advanced writer to a near-professional writer develops, the current study has, arguably, made a contribution to both SLA and language pedagogy. SLA research has only recently seen an increased interest in the development of advanced learners and studies that investigate advanced learners are few compared to those focusing on early or intermediate stages of development (Callies 2008). By describing the discourse competence of advanced Dutch learners of English this study has added new insights to this developing area of inquiry.

As the study focused on advanced learners writing academic English texts, the results of this study naturally have implications for EAP. First and foremost the results of the study contributes to the understanding of differences between native-speaker academic English and learner academic English, specifically Dutch learner academic English. The corpus study has confirmed some notions of Dutch learner language that experienced teaching professionals already suspected, for instance the underuse of non-finite clauses, disproved others, such as the underuse of non-restrictive relative clauses, uncovered some that had so far not been described in the literature, for instance the underuse of non-restrictive nominal appositions, and still others that are undetectable to the naked eye, but can only be uncovered through

quantitative analysis, for example the fact that although non-finite clauses are underused by learners they are only underused in medial and final positions in the sentence, not sentence initially. Moreover, this study has also proposed that these instances of overuse and underuse, regardless of their syntactic realization, can be grouped in systematic categories, which can be seen as dimensions in terms of discourse competence development. This means that future EAP pedagogical material can focus on these areas to help bridge the gap between what learners are capable of and the intended proficiency.

Other areas of EAP might also benefit from some of the insights gained from this study. The control corpus in the present study consists of writing from two academic disciplines, linguistics and literature. Many studies that look at academic writing use a corpus that consists of texts from many different academic disciplines, to discover features of academic writing that are common across academic disciplines (e.g. Coxhead 1998; Biber 1999). However, there is a growing number of EAP researchers (cf. Johns: 1997; Candlin and Plum: 1999; Hill 2005; Hyland 2002b, 2010, and Hyland & Hamp-Lyons 2002) who recommend that apart from teaching core features of academic writing it is also necessary to prepare learners for the specific literacy required in their respective disciplines. Having used a specialised corpus of literary and linguistic academic writing, as opposed to a 'generic' corpus of academic writing and highlighting some of the differences between the writing of these two disciplines as well as their similarities, this study also informs EAP in terms of genre specificity and contributes to the discussion regarding the level of specialization that is needed in EAP pedagogy.

Finally, Flowerdew (2001: 366) notes that very few findings of learner corpora research have been applied directly to pedagogy and seem to 'remain at the level of implications'. If ten years on there are editors that dare take up the challenge of developing EAP materials that address the specific problems non-native writers encounter, using empirical support from both native, genre-specific corpora and incorporating learner corpus research findings, they will find the four dimensions of discourse competence, put forward in this study, a useful guiding principle for the selection and ordering of pedagogical materials. This study has shown that with regard to discourse competence development Dimension 1 (concision), Dimension 2 (authorial confidence and involvement), and Dimension 3 (sophistication) are at least as relevant to advanced learners as the often-featured Dimension 4 (degree of formality). Using all four dimensions as a guide is a way of giving coherence to academic writing teaching methods.