Discourse connectives across CEFR-levels: A corpus based study

Cecilie Carlsen
University of Bergen

The chapter "Discourse connectives across CEFR-levels: A corpus based study" focuses on the use of discourse connectives, such as *and, but, so, then,* and *however,* in written learner texts of Norwegian as a second language. The *Common European Framework of Reference for Languages* (CEFR) makes specific predictions about the use of such discourse connectives in learner language, i.e. that that the range of different connectives expands across proficiency levels, that more advanced learners make use of less frequent connectives than learners at lower levels, and that learners gain increased control of connectives as they progress. The overall research question of the study reported in this chapter is whether the predictions made in the CEFR about learners' use of discourse connectives are supported by authentic learner data. The data used is a computer learner corpus of written Norwegian developed at the University of Bergen, Norway. This corpus has the great advantage of being linked to the CEFR. The study reported here is one small contribution to the huge task of validating the CEFR against real learner data, an overall aim of the SLATE network.

1. Introduction[1]

The present chapter focuses on the use of discourse connectives, such as *and, but, so, then,* and *however,* in written learner texts of Norwegian as a second language. The *Common European Framework of Reference for Languages*, CEFR, (Council of Europe, 2001) makes specific predictions about the use of such discourse connectives in learner language, elaborated in the illustrative scale of *Coherence and Cohesion* (p. 125). The CEFR predicts that the range of different connectives expands across proficiency levels, that more advanced learners make

---

[1] I would like to thank Daniel Apollon at the University of Bergen for invaluable help with the correspondence analysis, the editors and two anonymous reviewers for useful comments on earlier drafts of this chapter, and Tania Horak, at Lancaster University, for proof reading. Any remaining errors are my own.

use of less frequent connectives than learners at lower levels, and that learners gain increased control of connectives as they progress. The overall research question of the study reported here is whether the predictions made in the CEFR about learners' use of discourse connectives are supported by authentic learner data. The predictions are tested against a computer learner corpus of written Norwegian (ASK)[2] developed at the University of Bergen, Norway. This corpus has the great advantage of being linked to the CEFR, which allows us to investigate what learners can and cannot do at different CEFR-levels[3]. The study reported here is one small contribution to the huge task of validating the CEFR against real learner data, an overall aim of the SLATE network.

There have been relatively few studies of coherence in writing in Norwegian as a second language. One such study is Høyte (1997), who investigates the relation between learners' use of connectives and test-scores, and finds a weak positive correlation between scores and the use of varied connectives. Høyte does not, however, investigate further what connectives are used at low versus high levels of proficiency. In another study Palm (1997) compares native speakers and non-native speakers of Norwegian, and finds that the latter group over-uses the common connective *fordi* [because], but her study is based on a limited number of informants. Similarly, McGhie (2003) investigates the use of causal connectives in oral and written production of learners of Norwegian and finds that learners use only a few of the available connectives to express these rhetorical relations. Like Palm (2007) she finds an overuse of *fordi* [because] in learner language and a lesser use of its counterpart *derfor* [therefore]. Her study, however, includes only five informants. Qualitative studies are necessary in order to achieve an in-depth understanding of the semantic-pragmatic content and poly-functional use of connectives (Blakemore, 2002; Mosegaard-Hansen, 1998). The study presented in this chapter is however of a different kind: The approach is corpus-based and quantitative throughout. The research purpose is to look at the use of a range of different connectives in learner language at different levels of proficiency to reveal patterns of over- and underuse which may not be easily generalized from the results of studies based on small data samples. The study focuses on the use of 36 different connectives

----

2   ASK is an acronym for the three constituent morphemes of Norwegian AndreSpråksKorpus [SecondLanguageCorpus] (see Tenfjord, 2007, p. 207).

3   During 2008/2009 I collaborated with Felianka Kaftandjieva at the University of Sofia, Bulgaria, in linking ASK to the CEFR. A group of 10 experienced raters was involved in the re-assessment of corpus-texts. 200 texts were scored by 10 raters, the remaining 1022 texts by two parallel groups of 5 raters, each group scoring 511 texts (see Carlsen, 2010, for details).

in learner language at seven levels of proficiency and compared with the use of connectives of native speakers of Norwegian. This approach does not allow an in-depth investigation of the semantics and use of each connective[4]. It is the first study of connectives in Norwegian learner language using a learner corpus linked to the CEFR.

2. Theoretical background

2.1. Text coherence and discourse connectives

Coherence is often described as that: "[...] which makes a discourse more than the sum of the interpretations of the individual utterances" (Sanders & Spooren, 1999, p. 235). Coherence in a written text refers to the linking of ideas to make it meaningful to readers (Lee, 2002). The skilled writer uses a variety of devices to construct coherent texts such as reference, substitution and ellipsis (Halliday & Hasan, 1976; Kehler, 2004). The use of explicit linking words or linking phrases is one way of signalling coherence relations. A text may however be coherent without explicit marking of coherence relations, but such relations often are marked linguistically, as Knott and Dale (1994) and Spooren and Sanders (2008, p. 2005) point out.

The present study does not set out to investigate all available coherence devices, but limits its focus to explicit linking words such as *and, but, because, however, despite, furthermore*, referred to in the following as discourse connectives, or simply connectives (see Blakemore, 2002; Schiffrin, 1987). The class of connectives consists of different linguistic elements and is therefore difficult to define strictly grammatically. Most studies dealing with connectives therefore are confined to a functional definition (Mosegaard-Hansen, 1998). Fraser (1996, p. 190) defines connectives as elements "which signal a relation between the discourse segment which hosts them, and the prior discourse segment". I will use a similar definition here, but similarly to Spooren and Sanders (2008, p. 2005), I also include elements that link larger text sections and paragraphs. Discourse markers that do not primarily have a linking function, such as *I mean, sort of, right, well, oh, you know, kind of* etc., often called pause fillers or hesitation markers mostly used in spoken language, are not included in the present study.

---

4 This is the focus of a quantitative and qualitative follow-up study in which I look at the relative effect of proficiency-level and cross-linguistic influence on the use of a more limited number of connectives in the writing of Spanish learners of Norwegian.

2.2. Coherence and the use of connectives in learner language

Cross-linguistic studies comparing the use of discourse connectives between two or more languages have shown that different languages tend to use different connectives to a somewhat different degree and with somewhat different meaning (Fabricius-Hansen, 2005; Fløttum, Dahl, & Kinn, 2008; Östman, 2005; Stenstrøm, 2006). It is therefore not surprising that constructing coherent texts poses problems to language learners, even at advanced levels (Connor, 1996; Lee, 2002). Even so, the use of discourse connectives in learner language has not been the focus of much research interest (Hinkel, 2001, p. 113; Müller, 2005, p. 1). The existing research on the topic is hard to compare due to different definitions of connectives, and the results are inconclusive.

In research focusing on the relation between the use of connectives and levels of proficiency, some studies have found small or no differences between learners at different levels of proficiency (Castro, 2004; Johnson, 1992), while others have found that highly rated essays are cohesively denser than poorly rated ones (Witte & Faigley, 1981). Similarly, some early studies of learner English (Evensen, 1985; Rygh, 1986) found a higher frequency and diversity of connectors in texts produced by advanced learners than in texts by learners at lower levels of proficiency.

Other researchers have compared the use of connectives in native and non-native speaker texts. Connor (1984) found no significant difference in general cohesion density between native speakers and advanced learners, while others have found that non-native writers of English overuse explicit cohesion markers as compared to native English writers (Field & Oi, 1992).

Several researchers investigating the use of high-frequency connectives have come to the conclusion that these are overused by learners. In a corpus-based study, Paquot (2008) for instance compares the use of five exemplifying lexical items between non-native speakers (the International Corpus of Learner English) and native speakers (two different native corpora) and finds a striking overuse of the connective phrase *for example* by non-native speakers. Similarly, in a corpus-based study of Swedish learners of French (InterFra corpus) Hancock (2005) finds that the high-frequency connective *parce que* [because] is overused by learners even at advanced levels. Müller (2005), on the contrary, finds that native speakers of English use the simple causal marker *so* twice as much as non-native speakers.

Finally, Hinkel (2001) compares the use of coordinating conjunctions in texts written by native speakers of English and learner groups with different first languages (L1s), and finds that some L1-groups have a similar use to native speakers, some groups significantly underuse, while other groups overuse connectives, which points to the importance of studying discourse features in relation to cross-linguistic influence as well.

2.3. Coherence, connectives and the CEFR

In the CEFR discourse competence is treated as one aspect of pragmatic competence and defined as "the ability of a user/learner to arrange sentences in sequence so as to produce coherent stretches of language" (Council of Europe, 2001, p. 123). Coherence and cohesion are mentioned among other criteria which need to be met in order to achieve straightforward and efficient communication. Four illustrative scales are available for discourse competence (Council of Europe, 2001, pp. 124–125). In this study the focus is on the illustrative scale of *Coherence and Cohesion*, which describes the use of organisational patterns and cohesive devices in the construction of coherent discourse (see Appendix, Table A1 for a reproduction of the illustrative scale of coherence and cohesion).

The scale of coherence and cohesion reflects a basic distinction in the description of language development in the CEFR, i.e. that between quantity and quality. The use of connectives across proficiency levels is, on the one hand, described in relation to the relative range of different connectives used: At the lower levels, only the "very basic connectors" are expected, "simple connectors" are expected at the A2-level and at the A2+ level "the most frequently occurring connectors". At the higher levels, the range of different connective devices is assumed to increase. It is worth noticing that it is only at the B2+ level that "a variety of linking words" is expected. Below B2+ level only "a limited number of cohesive devices" are expected. At this point, an important distinction needs to be made very clear: The CEFR predicts greater range but does *not* predict greater connective density at higher levels. It does not predict that the more advanced the learners get, the more overt signals of coherence relations they use. Since the main purpose of the study reported here has been to test the predictions of the CEFR, connective density has not been the focus of my study.

On the other hand, the CEFR describes the use of connectives in relation to the degree of control and efficiency with which they are employed: At the A1, A2 and B1 levels limited reference is made regarding the control of connectives, other than "can link…". At the B2-level, connectives are described as linking utterances into a "clear, coherent discourse, though there may be some "jumpiness" in a long contribution". At the B2+ level, there is explicit reference to the use of connectives as being "efficient", at C1 as being "controlled" and finally at the C2-level: "full and appropriate".

3. Aim of the study

The overall research question of the study reported here is whether the CEFR's description of learners' use of connectives is supported by empirical learner data. Based on the level descriptors of the illustrative scale of coherence and cohesion, three main predictions about the use of connectives may be deduced. Firstly, the *range* or repertoire of cohesive devices is assumed to grow across CEFR-levels. If the scale's predictions are correct, learners at higher levels of proficiency should utilise a greater variety of different connectives than learners at lower levels. Secondly, learners at lower levels of proficiency are assumed to rely heavily on the use of common, high-frequency connectives, while learners at higher levels are assumed to use *low-frequency* connectives as well. And finally, learners at higher levels of proficiency show qualitatively better *control* of cohesive devices than learners at lower levels[5]. The above predictions have been reformulated into three hypotheses, which are tested empirically in this study:

> H1: Texts at higher levels contain a *broader range* of different cohesive devices than texts at lower levels
>
> H2: Texts at higher levels contain more *low-frequency* connectives than texts at lower levels
>
> H3: Texts at higher levels show a *greater degree of control* of the cohesive devices used than texts at lower levels

4. Data and methodology

In order to investigate the predictions made in the CEFR about the use of connectives, it was necessary to include a wide range of different connectives in the study. Since no definitive list of discourse markers exists (Blakemore, 2004, p. 221), I had to develop a list based on earlier taxonomies of connectives (Halliday & Hasan, 1976; Knott & Dale, 1994) and descriptions in Norwegian grammars (Faarlund, Lie, & Vannebo, 1997; Hagen, 1998). I wanted to include a range of different connectives representing different rhetorical functions as

---

5   A fourth prediction could also be made; i.e. the reference to specific connectives (*and, then, but, because*) at the A1- and A2-levels. Since the target language is Norwegian and not English, and since there are no A1-texts in the ASK-corpus at present, this prediction has not been tested here.

mentioned in Halliday and Hasan's (1976, pp. 242–243)[6] taxonomy, e.g. *additive, adversative* and *causal*, single- as well as multi-word units, and connectives with different degrees of frequency. The frequency of the different connectives was established by investigating their use in texts written by native speakers of Norwegian selected from the control corpus of ASK, which contains texts written by 200 native speakers of Norwegian on the same prompts and under the same circumstances as the other texts in ASK (native speaker (NS)-ASK). Connectives that were not used in the NS-ASK were excluded from investigation. The connectives were divided into three frequency groups: High-frequency, n = 5 (relative frequency > 1.0), Medium-frequency, n = 19 (relative frequency < 0.9999 and > 0.0100), and Low-frequency, n = 12 (relative frequency < 0.0100). The total number of connectives included in this study is 36 (see Table 1 below for the categorization and translation of the Norwegian connectives into English[7]).

The data used in this study are selected from the electronic learner corpus of Norwegian (Norsk Andrespråkskorpus, ASK, see Tenfjord, 2007). ASK contains texts written by learners of Norwegian with 10 different L1s (Albanian, Bosnian/Serbian/Croatian, Dutch, English, German, Polish, Russian, Spanish, Somali and Vietnamese). The informants are adult foreigners living in Norway, making the corpus one of second and not foreign language. The texts are authentic test responses (essays) selected from two different standardized tests of Norwegian as a second language, one at the intermediate level and one at the advanced level. At the intermediate level, learners are asked to "write a text" about everyday themes, for example traditions, values, friendship, the place you live etc. At the advanced level, learners are asked to discuss and develop an argument in relation to themes such as education, integration, welfare, pollution, labour etc. The text types or genres are somewhat different at the two levels, mainly descriptive/expository at the intermediate level and expository/argumentative at the advanced level. The different genres are therefore likely to be reflected in the rhetorical function of the connectives used: More adversative

---

6 Halliday and Hasan's taxonomy also includes temporal connectives, which has been largely ignored in my study: Many of them have non-connective homonyms in Norwegian, which makes them notoriously difficult to study in a quantitative study such as the present one. Secondly, distinguishing between items referring to external events and connectives referring to the order of the text itself requires a qualitative approach.

7 The translation is tentative and merely linguistic since different languages use connectives in somewhat different ways, as cross-linguistic studies referred to here show.

and causal connectives are to be expected at the higher levels of proficiency. Genres should not, however, affect the hypotheses tested in this study since the use of varied connectives, low-frequency connectives, and control of connectives may be observed within each rhetorical group of connectives.

The corpus texts are automatically tagged for word class and morphological traits and manually error-coded, which allows searches of words and word-combinations, of incorrect as well as of correct forms. During the year of 2008/2009 two thirds of the texts in ASK (1222 texts) representing all texts of 7 L1s were subject to a re-assessment on the CEFR-scale by a group of 5-10 experienced raters who are very familiar with the CEFR levels (Carlsen, 2010). A series of different reliability estimates were calculated, such as *Homogeneity index* (Mean +0.84), *Correlation with the rest* (+0.90), and *Inter-rater correlation* (Mean +0.82), all well within an acceptable range in terms of rater agreement. Whole levels (A2, B1, B2, and C1) as well as in-between levels (A2/B1, B1/B2, and B2/C1) were used[8]. The size of the different CEFR-groups varies from 137,885 words in the B2-group to 6,115 words in the A2-group (see Appendix, Table A2 for CEFR-group size).

The study reported in this chapter uses a quantitative method and is to a large degree based on investigations of frequency of use. The mentioned difference in size of the CEFR-groups makes it necessary to use relative rather than absolute frequencies. Relative frequencies are calculated automatically in ASK by dividing the absolute frequency of occurrences by the number of words within each level group. The occurrences of connectives across the CEFR-levels (see H1 and H2) were investigated through the means of *correspondence analysis*, which is an exploratory technique designed to analyse the relation between rows and columns in a two- or multi-way table. The results are usually displayed as a scatter plot, in which the relations between, for example, variables on the one hand and observations on the other are visualized jointly as points in a common coordinate system. If the original data set is high dimensional, the reduction obtained in a much lower dimensional space may offer substantial advantages for interpreting the latent structure of the data set. Correspondence analysis is therefore a useful tool to get an overview of patterns in a data set.

The final hypothesis, H3, should ideally have been investigated qualitatively. This has not been practically possible due to the large number of connectives included. I have therefore used a quantitative approach based on the error-cod-

---

8   There were no texts in ASK found to be at levels below A2 or above C1. ASK is however currently being expanded by adding more texts at the A1/A2, A2, and A2/B1 levels.

ing inherent in ASK. Based on the results of the correspondence analysis, there are good reasons to ignore the high frequency connectives, which, as we will see, are used similarly across CEFR-levels, and focus attention on the medium- and low-frequency connectives. To test H3, only connectives that are actively used by all groups have been included, since control of use cannot be observed unless there are actual occurrences in the learner texts. Error searches were restricted to *W- wrong word choice,* while errors in spelling, morphology or syntax were ignored. The results are presented as errors across the number of total occurrences and as relative frequencies of errors, calculated by dividing the number of errors by the number of total occurrences within each CEFR-group (see Appendix, Table A3).

5. Analysis and interpretation

The results of the correspondence analysis are presented in Figure 1 below[9]. Since connectives are used somewhat differently across different languages, a translation of the connectives in the figure to English is not straightforward, even though it would make the figure more understandable to the general reader. I have therefore chosen to keep the Norwegian labels in the figure, making a translation available in Table 1 presented below. The table also explains the abbreviations of some of the connective phrases necessary in order to make Figure 1 more readable. The most important information in relation to the CEFR and the hypotheses of this study is not on the level of the individual connectives, but rather on the group-level based on frequency: H (high-frequency), M (medium-frequency) and L (low-frequency), added in front of each connective in Figure 1.

The scatter plot in Figure 1 displays the relation between the two variables of the dataset, CEFR-levels and connectives. The CEFR-levels order the use of connectives along a gradient stretching roughly from east (lowest proficiency values) to west (highest proficiency values). The vertical line running through the barycentre (centre of mass, 0) between B1 and B1/B2 discriminates efficiently between patterns of use typical of lower level groups (A2 to B1) and higher level groups (B1/B2 to C1). The clustering of upper level groups (B1/B2-C1) indicates only minor differences within these groups when it comes to the use of connectives. The more advanced learners' use of connectives is sim-

---

9 The most frequently used connective by all groups, *og* [and], is not included in the figure. It is used similarly across proficiency groups.

Figure 1. Results of the correspondence analysis

ilar to that of the native speakers (NS-ASK), with the exception of three con-
nectives located in the upper-left corner of the scatter plot. Connectives near the
barycentre, such as *men* [but], *eller* [or], and *også* [too/also] are common to all
groups and do not differentiate between higher and lower levels. 75.52% of the
variance (inertia) of the original data set is explained by two dimensions (axis 1:
61.96%, axis 2: 13.57%) indicating a high degree of correlation between
CEFR-levels and the use of connectives, and a clearly different use of connec-
tives at lower versus higher levels of proficiency.

The prediction of H1 was that texts at higher levels would contain a *broad-*
er range of different cohesive devices than texts at lower levels. This prediction
is supported by the data. The scatter plot demonstrates that a great number of
different connectives clustered around the higher CEFR-levels to the left in the

Table 1. Connectives in frequency-groups based on NS' use of connectives, with translation into English

FUNCTION	HIGH-FREQ.		MEDIUM-FREQUENCY		LOW-FREQUENCY	
ADDITIVE	*eller* *også*	or too/also	*for eksempel* *i tillegg* *enten..eller* *ikke minst* *heller ikke* *først og fremst* *for det første (fdf)* *dessuten*	for example in addition either….or not least nor first and foremost firstly besides	*det vil si* *med andre ord* *for det andre (fda)* *bortsett fra* *det betyr*	i.e./this means in other words secondly except from this means
ADVERSATIVE	*men*	but	*likevel* *selv om* *derimot* *imidlertid*	still/nevertheless even though on the other hand however	*istedenfor* *til tross for (ttf)* *dog* *på den ene siden* *på den andre siden* *ikke desto mindre*	instead of despite though on the one hand on the other hand nevertheless
CAUSAL	*fordi*	because	*derfor* *slik at* *så (saa)* *på grunn av* *siden* *ettersom* *for*	therefore so that so because of since since for	*følgelig*	consequently

picture, and only a few to the right, which indicates that learners at higher levels of proficiency use a series of different connectives, while the lower-level learners confine themselves mainly to just a few connectives. One important point should be made: The CEFR predicts that even at a B2-level only a limited number of cohesive devices are used. The results of the present study clearly show, however, that learners even at a B1/B2 level use a range of different connectives, and to an extent which separates them sharply from the lower levels.

H2, which implies that learners at higher levels of proficiency use more *low-frequency* connectives than learners at lower levels, is also supported by the data. Around the lower level groups on the right side of the vertical line, there are mainly high- and medium-frequency connectives, while there are a great number of low-frequency connectives clustered around the more advanced learners to the left. The high-frequency connectives, except H-*fordi* [because], and H-*også* [too/also], are all clustered around the barycentre and common to all groups. Low-frequency connectives are generally under-used at A2, A2/B1

and B1-levels, with the exception of the low-frequency connective phrase L-*det betyr* [this means] which is overused at the A2-level.

According to the final hypothesis, H3, learners at higher levels of proficiency show a *greater degree of control* of the cohesive devices used (see Appendix, Table A3). The results of the error-searches show that there are in fact relatively few cases where learners use connectives wrongly, in the sense that they use one connective where they should have used another one instead. However, the relative frequency of errors is strongly correlated with CEFR-level (Chi-square test: p<0.001 level), yielding support for the prediction made in the CEFR that the degree of control of cohesive devices rises across proficiency levels. An interesting point worth mentioning here is the fact that the error-pattern differs across the rhetorical function of the connectives: The additive connectives are used correctly in all instances by the B2, B2/C1 and C1 groups, and also in the groups of A2 and A2/B1, and there are only occasional errors in the B1 and B1/B2 groups. The adversative and causal connectives on the other hand, have a somewhat higher error rate. The connectives that seem to cause the most problems to learners in my study are the adversative connectives *derimot* [however/contrary] and *selv om* [even though], as well as the causative connective *derfor* [therefore].

In this study the main focus has not been on the individual connectives, yet the study has shed some light on the use of particular connectives across CEFR-levels. Most of the 36 connectives can be divided into three groups according to the pattern they show across the CEFR-levels: Some connectors show a steep negative correlation across levels of proficiency. They are largely over-used at low levels of proficiency and fall gradually as one approaches the C1-level. This is the case for the high-frequency connectives *men* [but], and *fordi* [because] (the latter is used six times as frequently by the A2-group as by the C1 group), for the medium-frequency connectives *så* [so], and *for* [because], as well as for the low-frequency connective phrase *det betyr* [this means] (see Appendix, Figures A1 and A2). Other connectives show an opposite pattern, i.e. they are only used to a limited degree or not used at all at low levels of proficiency and increase with higher levels of proficiency. This is the case for a range of medium-frequency connectives, for example *i tillegg* [in addition], *slik at* [so that], *ikke minst* [not least], *derfor* [therefore] and *enten…eller* [either…or] (see Appendix, Figures A3 and A4). Finally, quite a few connectives increase in use from A2 to B2 - B2/C1 levels and then drop as one approaches the C1-level. Many low-frequency connectives show this pattern, for example *istedenfor* [instead of]*, på den ene siden* [on the one hand], *på den andre siden* [on the other hand], *ikke desto mindre* [nevertheless], *til tross for* [despite], and *med andre ord* [in other words] (see Appendix, Figures A5 and A6).

6. Discussion and conclusions

The results of the study reported here are not surprising as they largely support the predictions made in the CEFR. Still, some of the findings may need further comment. As predicted by H1, learners at higher levels of proficiency tend to use a greater range of different connectives in their writing than learners at lower levels. This finding is in line with Evensen (1985) and Rygh (1986). The data of the present study do however show that learners use a range of different connectives earlier than predicted in the CEFR: Learners even at B1/B2-level use a range of different connectives contrary to the predictions of the CEFR, which only expects this at the B2 + level. This finding may indicate that a revision of the CEFR-scale is warranted at this point.

H2 was also largely supported by the data. Indeed, learners at lower levels tend to overuse high-frequency connectives, like *fordi* [because], *men* [but], *eller* [or] and medium-frequency connectives like *så* [so], and *for* [because]. This phenomenon is well-known from other studies of connectives (Hancock, 2005; Paquot, 2008). It is tempting to borrow a term from a former colleague and associate of the SLATE-network, Angela Hasselgreen, who refers to this phenomenon in learners' use of vocabulary as "lexical teddy bears" (Hasselgreen, 1994). High frequency connectives or "connective teddy bears" give novice learners a degree of comfort and security particularly useful in a test-situation like the one from which the texts in the ASK-corpus are selected. It is likely that the overuse of high-frequency connectives is due in part to learners using one and the same connective expressing different rhetorical functions. The obvious limitation of a quantitative study as the one presented here is the lack of information about the use and content of the different connectives in learner language, and it is therefore important to complement studies like the present one with qualitative investigations.

The data also support H3 indicating that learners gain increased control of the connectives they use across levels of proficiency. There are only a few mistakes in their use but still, the lower-level groups make significantly more errors than advanced learners. The data of the present study do not lend themselves to investigating the reasons why there are more errors in the use of adversative and causal connectives than in the use of additive connectives. The rhetorical functions expressed by adversative and causal connectives are more complex than a mere adding of information, which may explain some of the difference. In addition, the adversative and causal connectives included in the study of H3 have more specific content and use, which makes errors easier to spot. Again, these questions need to be addressed empirically in a qualitative study.

Finally, the results of the study indicate that some low-frequency connectives are used gradually more frequently towards higher levels of CEFR, but

drop at the highest levels. A possible explanation may be that while learners at B2 and B2/C1 levels use explicit low-frequency markers consciously when structuring their texts, learners at C1 level use other devices to express coherence. As mentioned earlier, the use of explicit coherence markers like connectives is but one way of constructing coherent texts. The mere presence of connectives does not necessarily make a text coherent, and indeed, it is possible to construct a coherent text with limited use of explicit markers of coherence relations. The finding illustrates that coherence-relations need to be investigated qualitatively as well as quantitatively to grasp the full picture.

References

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse connectives.* Cambridge: Cambridge University Press.

Blakemore, D. (2004). Discourse markers. In R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 221–240). Malden: Blackwell.

Carlsen, C. (2010). Linking a learner corpus to the Common European Framework of Reference. Manuscript submitted for publication.

Castro, C. D. (2004). Cohesion and the social construction of meaning in the essays of Filipino college students' writings in L2 English. *Asia Pacific Education Review, 5,* 215–225.

Connor, U. (1984). A study of cohesion and coherence in English as a second language students' writing. *Research on Language & Social Interaction, 17,* 301–316.

Connor, U. (1996). *Contrastive rhetoric. Cross-cultural aspects of second-language writing.* Cambridge: Cambridge University Press.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Cambridge: Cambridge University Press.

Evensen, L. S. (1985). Discourse-level interlanguage studies. In N. E. Enkvist (Ed.), *Coherence and composition. A symposium* (pp. 39–65). Åbo: Åbo Akademi.

Faarlund, J. T., Lie, S., & Vannebo, K. I. (1997). *Norsk referansegrammatikk.* Oslo: Universitetsforlaget.

Fabricius-Hansen, C. (2005). Elusive connectives. A case study on the explicitness dimension of discourse coherence. *Linguistics, 43,* 17–48.

Field, Y., & Oi, Y. L. M. (1992). A comparison of internal conjunctive cohesion in English essay writing of Cantonese speakers and native speakers of English. *RELC Journal, 23*(1), 15–28.

Fløttum, K., Dahl, T., & Kinn, T. (2008). *Academic voices.* Amsterdam: John Benjamins.

Fraser, B. (1996). Pragmatic markers. *Pragmatics 6*(2), 167–190.

Hagen, J. E. (1998). *Norsk grammatikk for andrespråkslærere.* Oslo: Ad Notam Gyldendal.

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

Hancock, V. (2005). Étude intonodiscursive des sequences introduites par parce que dans le francais parlé des apprenant avancés suedophones. In J. M. Delefosse (Ed.), *Actes du colloque international organisé les vendredi 25 et samedi 26 juin 2004 à Paris. Acquisition, pratiques langagières, interactions et contacts (APLIC)* [CD-ROM]. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-10149

Hasselgren, A. M. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with vocabulary. *International Journal of Applied Linguistics, 4,* 237–258.

Hinkel, E. (2001). Matters of cohesion in L2 academic texts. *Applied Language Learning, 12,* 111–132.

Høyte, T. (1997). *Tekstkoherens og tekstkvalitet* (Unpublished master's thesis). University of Bergen, Norway.

Johnson, P. (1992). Cohesion and coherence in compositions in Malay and English. *RELC Journal, 23*(2), 1–17.

Kehler, A. (2004). Discourse coherence. In L. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 241–265). Malden: Blackwell.

Knott, A., & Dale, R. (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes, 18,* 35–62.

Lee, I. (2002). Teaching coherence to ESL students: A classroom inquiry. *Journal of Second Language Writing, 11*(2), 135–139.

McGhie, M. (2003). Årsaksrelasjoner i tilegnelse av norsk som andrespråk. In S. Lie, G. Nedrelid, & H. Omdal (Eds.), *MONS 10. Utvalde artiklar frå det tiande møte om norsk språk* (pp. 207–217). Kristiansand: Høyskoleforlaget.

Mosegaard-Hansen, M.-B. (1998). The semantic status of discourse markers. *Lingua, 104,* 235–260.

Müller, S. (2005). *Discourse markers in native and non-native English discourse.* Amsterdam: John Benjamins.

Östman, J.-O. (2005). Constructions in cross-language research. Verbs as pragmatic particles in Solv. In K. Aijmer & A.-M. Simon-Vandenbergen (Eds.), *Pragmatic markers in contrast* (pp. 237–257). Amsterdam: Elsevier.

Palm, K. (1997). *Argumenterende skriving. En analyse av andrespråkstekster og førstespråk-stekster fra videregående skole* (Unpublished master's thesis). University of Oslo, Norway.

Paquot, M. (2008). Exemplification in learner writing: A cross-linguistic perspective. In F. Meunier & S. Granger (Eds.), *Phraseology. An interdisciplinar perspective.* Amsterdam: John Benjamins.

Rygh, I. L. (1986). Connector density? an indicator of essay quality? In L. S. Evensen (Ed.), *Nordic research in text linguistics and discourse analysis* (pp. 203–213). Trondheim: Tapir.

Sanders, T., & Spooren, W. (1999). Communicative intentions and coherence relations. In W. Bublitz, M. Lenz, & E. Ventola (Eds.), *Coherence in spoken and written discourse* (pp. 235–350). Amsterdam: John Benjamins.

Schiffrin, D. (1987). *Discourse markers.* Cambridge: Cambridge University Press.

Spooren, W., & Sanders, T. (2008). The acquisition order of coherence relations: On cognitive complexity in discourse. *Journal of Pragmatics, 40,* 2003–2026.

Stenström, A.-B. (2006). The Spanish discourse markers o sea and pues and their English correspondences. In K. Aijmer & A.-M. Simon-Vandenbergen (Eds.), *Pragmatic markers in contrast* (pp. 155–172). Amsterdam: Elsevier.

Tenfjord, K. (2007). ASK and you will find what you seek. In C. Carlsen & E. Moe (Eds.), *A human touch to language testing* (pp. 198–208). Oslo: Novus Press.

Witte, S. P., & Faigley, L. (1981). Coherence, cohesion and writing quality. *College Composition and Communication, 32,* 189–204.

APPENDIX

Table A1. The illustrative scale of Coherence and Cohesion (Council of Europe, p. 125).

	COHERENCE AND COHESION
C2	Can create coherent and cohesive text making *full and appropriate use* of a variety of organisational patterns and a *wide range of cohesive devices.*
C1	Can produce clear, smoothly flowing, well-structured speech, showing *controlled use* of organisational patterns, *connectors and cohesive devices.*
B2	Can use a *variety of linking words efficiently* to mark clearly the relationships between ideas.
	Can use a *limited number of cohesive devices* to link his/her utterances into clear, coherent discourse, though there may be some "jumpiness" in a long contribution.
B1	*Can link* a series of shorter, discrete simple elements into a connected, linear sequence of points.
A2	Can use *the most frequently occurring connectors* to link simple sentences in order to tell a story or describe something as a simple list of points.
	Can link groups of words with *simple connectors* like "and", "but" and "because".
A1	Can link words or groups of words with *the very basic linear connectors* like "and" or "then".

Table A2. Number of words in the different CEFR-groups in ASK

| Number of words	CEFR-level
6137	A2
52221	A2/B1
92334	B1
89617	B1/B2
137413	B2
41383	B2/C1
11614	C1

Table A3. Erroneous use of a sample of the 36 connectives across CEFR-levels*

| FUNCTION	L- and M-freq. connectives used by all groups	A2, A2/B1	B1, B1/B2	B2, B2/C1, C1
ADD.	M-for eksempel/ **for example**	0/56	1/177	0/181
	[rel.freq.]	[0]	[0,0056]	[0]
	M-i tillegg/ **in addition**	0/11	0/58	0/116
	[rel.freq.]	[0]	[0]	[0]
	M-først og fremst/ **first and foremost**	0/16	1/81	0/93
	[rel.freq.]	[0]	[0,0123]	[0]
	M-dessuten/ **besides**	0/15	1/77	0/79
	[rel.freq.]	[0]	[0,0129]	[0]
	L-det vil si/ **this means**	0/5	0/25	0/41
	[rel.freq.]	[0]	[0]	[0]
	L-det betyr/ **this means**	0/14	0/65	0/61
	[rel.freq.]	[0]	[0]	[0]
ADV.	M-likevel/ **still, however**	1/17	1/59	0/102
	[rel.freq.]	[0,0588]	[0,0169]	[0]
	M-selv om/ **even though**	2/31	2/134	2/167
	[rel.freq.]	[0,0645]	[0,0149]	[0,0119]
	M-derimot/**on the other hand**	1/1	1/9	2/37
	[rel.freq.]	[1.0000]	[0,1111]	[0,0540]
	L-til tross for/ **despite**	0/1	0/15	0/15
	[rel.freq.]	[0]	[0]	[0]
CAUS.	M-derfor/ **therefore**	3/83	1/246	1/272
	[rel.freq.]	[0,0361]	[0,0040]	[0,0036]
	M-på grunn av/ **because of**	0/33	1/141	0/123
	[rel.freq.]	[0]	[0,0070]	[0]
	M-ettersom/ **since**	0/2	0/1	0/4
	[0]	[0]	[0]	[0]
ERRORS TOTAL				
REL. FREQ.		**1,1594**	**0,1846**	**0,0695**

* The total number of occurrences is presented to the right of the slash, while the number of errors in use of the particular connective is presented to the left, for example *selv om* [even though] is used 31 times in the lowest proficiency group, 134 times in the B1 and B1/B2 groups and 167 times by the most proficient learners. In each proficiency group there are two occurrences where the connective is used incorrectly in the sense that a different connective should have been chosen instead. The number of errors (wrong choice of connective) is divided by total occurrences of the particular connective within the CEFR-level group to obtain the relative frequency of errors. Since the connective *selv om* is used five times as often in the texts of the more advanced learners, the relative frequency of errors is smaller [0.0119] in this group than in the lowest proficiency group [0.0645]. The last row displays the total occurrence of errors in each group divided by the actual occurrences within each group, i.e. it is the sum of the relative frequencies for each group.

Even though the number of errors is not high for any group, there is a positive correlation between levels of proficiency and control when defined as the lack of errors in use. The difference between level groups for total errors is significant at the p>0.001 level (Chi-square "Goodness of Fit" Test).

Figures A1 and A2: Connectives overused at lower levels (NS-ASK to the right)
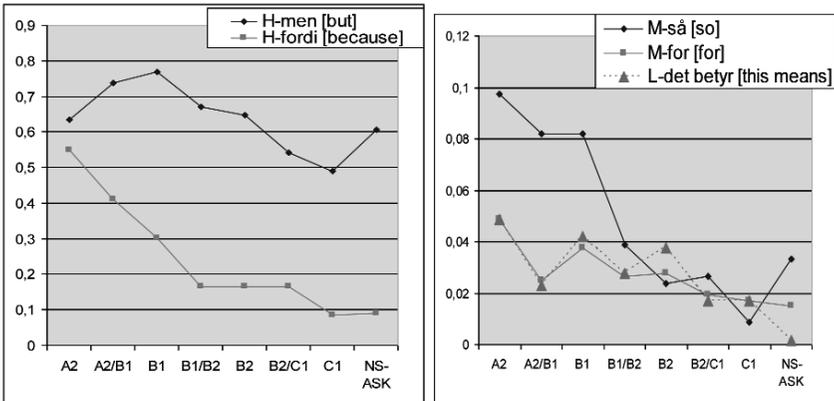
Figures A3 and A4: Connectives underused at lower levels (NS-ASK to the right)
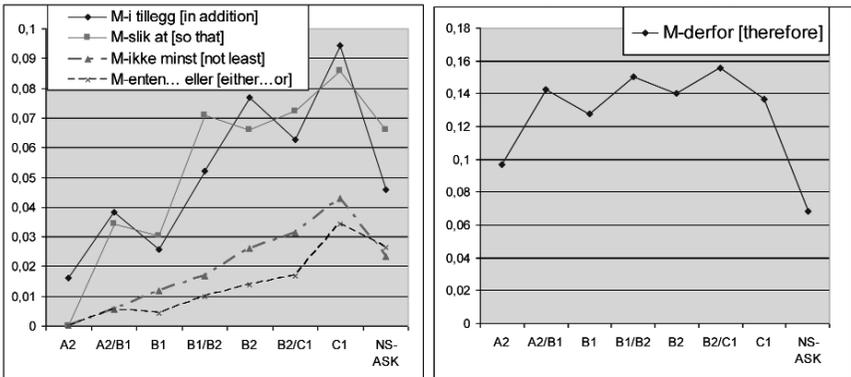
Figures A5 and A6: Connectives rise in use, drop at B2-B2/C1 levels (NS-ASK to the right)