At the interface between Contrastive Analysis and Learner Corpus Research: A parallel contrastive approach[1]

Hilde Hasselgård and Signe Oksefjell Ebeling, University of Oslo

Abstract
This paper presents a model for combining contrastive analysis and interlanguage analysis. It can be seen as an extension of Granger's (1996) Integrated Contrastive Model, but it explicitly requires a bidirectional parallel corpus for the contrastive analysis and matching learner corpora for the (multiple) interlaguage analysis. Thus, the model operates at the interface between Contrastive Analysis and Learner Corpus Research. Within the model, the contrastive, cross-linguistic analysis produces hypotheses for learner behaviour in the languages examined. Hence the learner corpus study of L2 data includes a cross-linguistic interlanguage analysis, based on comparable data produced by (at least) two learner groups (with different L1s and different L2s). As an illustration of the model, the English noun *people* and its typical Norwegian translation correspondences (FOLK and MENNESKE) are studied, with particular attention to postmodification patterns. The proposed *Parallel Contrastive Model* is shown to have the potential to throw new light on the cross-linguistic relationship between languages and interlanguages in a common framework.

Keywords: contrastive analysis, learner corpus research, noun modification, English/Norwegian, Parallel Contrastive Model, parallel interlanguage analysis

1. Introduction

Contrastive analysis and interlanguage analysis are both fields of research in their own right. However, there are obvious connections between them, as evidenced for example in Lado's much-quoted view that "in the comparison between native and foreign language lies the key to ease or difficulty in foreign language learning" (1957: 1). This has later been referred to as the "contrastive analysis hypothesis" (Wardhaugh 1970). In its strong version, presuming that cross-linguistic differences can reliably predict or diagnose problems in second language

---

[1] A version of this paper was given at a pre-conference workshop in honour of Sylviane Granger entitled "LCR at the interfaces", hosted by the 4th Learner Corpus Research conference in Bozen/Bolzano (LCR2017).

learning, Wardhaugh considers the hypothesis to be "quite unrealistic and impracticable" (ibid.: 124). The weak version, however, is seen as a more helpful perspective as it "requires of the linguist only that he use the best linguistic knowledge available to him in order to account for observed difficulties in second language learning" (ibid.: 126). This is also akin to the principle underlying Granger's (1996) *Integrated Contrastive Model* (ICM), which "involves constant to-ing and fro-ing between CA and CIA. CA data helps analysts to formulate predictions about interlanguage which can be checked against CIA data" (ibid.: 46). By CA is meant Contrastive Analysis between languages, i.e. cross-linguistic analysis, and by CIA is meant Contrastive Interlanguage Analysis, i.e. analysis of the language produced by a group of learners of a second or foreign language in comparison with native speakers and/or another group of learners.

In recent years, corpora and corpus-linguistic methods have had a great impact on the modern versions of contrastive and interlanguage analysis, to the extent that we can now talk about the fields of Corpus-based Contrastive Analysis and Learner Corpus Research (LCR). Although LCR does not by definition require a contrastive dimension as an integral part of the interlanguage analysis, "CIA has become a highly popular methodological approach within LCR" (Granger 2015: 19).

This paper explores a way of strengthening the connection between corpus-based contrastive analysis and interlanguage analysis by adding a cross-linguistic analysis of learner languages. That is, we juxtapose a contrastive analysis, based on a bidirectional parallel corpus, following the principles of Johansson's bidirectional model (Johansson & Hofland 1994; Johansson 2007), with the analysis of comparable L2 data in both of the languages contrasted, in line with Granger's ICM; see further below. The proposed *Parallel Contrastive Model* thus constitutes an extension of the Integrated Contrastive Model (Granger 1996; Gilquin 2000/2001) in that the contrastive analysis forms the basis for hypotheses about learner behaviour in both of the languages explored and in that it enables a parallel interlanguage analysis. The Parallel Contrastive Model will be illustrated by a study across languages and interlanguages. The aim of this exercise is, first and foremost, to give a practical demonstration of the model and to explore how it may shed some light on how learners of two different languages handle the same linguistic phenomenon, and on the extent to which the similarities and differences

attested in comparable L1 varieties are also evident in comparable L2 varieties. In other words, we seek to demonstrate how insights from cross-linguistic analysis can be applied to contrastive analysis between interlanguage varieties of the same pair of languages. In our case, the language pair is English and Norwegian, studied contrastively through both L1 and L2 corpora (see further Section 2.1).

The paper is organized as follows: Section 2 introduces the Parallel Contrastive Model against the backdrop of two existing models. Some previous research relevant to the model is outlined in Section 2.2, after which an overview of the corpora needed to illustrate it is given in Section 2.3. Section 3 offers a detailed description of the methodological steps prompted by the design of the model, exemplified through a parallel contrastive analysis of PEOPLE-nouns and their modification patterns in English and Norwegian L1 and L2.[2] Section 4 offers some further discussion and concluding remarks.

2. Linking up Contrastive Analysis and Learner Corpus Research

As the title of this paper suggests, the main focus of this paper is on the interfaces between Contrastive Analysis and Learner Corpus Research, as we believe that the potential of combining the two has not yet been explored fully. In an attempt to investigate this potential a bit further, the following two relevant models will be merged: Johansson's parallel corpus model (Johansson & Hofland 1994) and Granger's Integrated Contrastive Model (1996). The resulting Parallel Contrastive Model is outlined below, along with its potential for extending the scope of both contrastive and interlanguage analysis based on corpora.

2.1 Introducing the Parallel Contrastive Model

In the Integrated Contrastive Model (ICM), Contrastive Analysis (CA) between languages is combined with a contrastive analysis of interlanguages, also known as Contrastive Interlanguage Analysis (CIA). Figure 1 shows Granger's original representation of the model (Granger 1996: 47), in which CIA is typically carried out by contrasting

---

[2] 'PEOPLE-nouns' is used as a collective reference to the lemmas PEOPLE, FOLK and MENNESKE.

interlanguage with native language (IL < > NL),[3] or interlanguage with interlanguage (IL < > IL), for example the English produced by learners with different L1 backgrounds. In the upper half of the model, CA is shown to have two branches, involving CA on the basis of either comparable original data in two languages (OL < > OL), or translation data between two languages (SL < > TL).[4] From an interlanguage research perspective, the CA part of the model has the potential of either predicting or diagnosing transfer from the learner's L1.

Although Johansson's parallel corpus model has much in common with the CA part of the ICM, it is probably more robust in that it explicitly requires a bidirectional translation corpus, enabling CA on the basis of comparable and translation data within the same model. This is important for a number of reasons that we will return to below (steps 1 and 2 in the Parallel Contrastive Model).
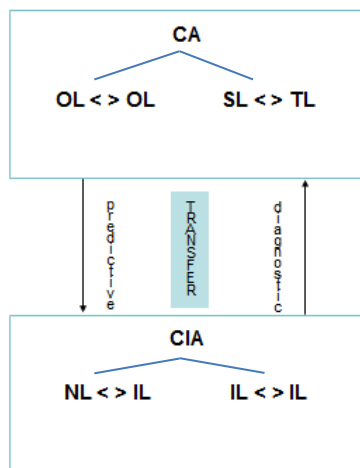

Figure 1 The Integrated Contrastive Model (Granger 1996: 47)

The parallel corpus model was first devised for the English-Norwegian Parallel Corpus; thus, Figure 2 includes reference to these languages. Figure 2 presents an adapted version of Johansson's parallel

---

[3] NL, or native language, is now often referred to as "reference variety" (Granger 2015).
[4] OL = original language; SL = source language; TL = target language.

corpus model where only the types of comparisons that are relevant for the ICM are shown, i.e. English originals vs. Norwegian originals (slant line), English originals vs. Norwegian translations and Norwegian originals vs. English translations (horizontal lines). Lines pointing to potential comparisons between e.g. English originals and English translations or between Norwegian translations and English translations are therefore left out here.
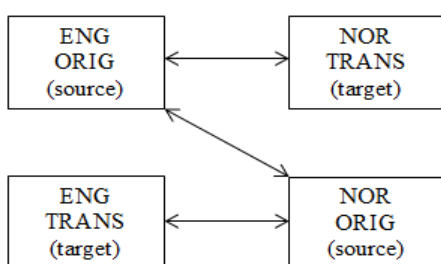

Figure 2 Johansson's parallel corpus model (adapted from Johansson & Hofland 1994)

In the suggested Parallel Contrastive Model we explicitly bring Johansson's parallel corpus model into the CA part of Granger's Integrated Contrastive Model, while in the CIA part we suggest a combination of Contrastive Interlanguage Analysis (CIA) and Parallel Interlanguage Analysis (PIA), the latter by analogy of the parallel corpus model in which comparable data in two L2s, rather than two L1s, are compared and contrasted. Figure 3 outlines the model with references to the corpora and languages that will be used for illustration (see Section 3). A detailed account of the corpora – ENPC, ICLE, and ASK – will be given in Section 2.3.
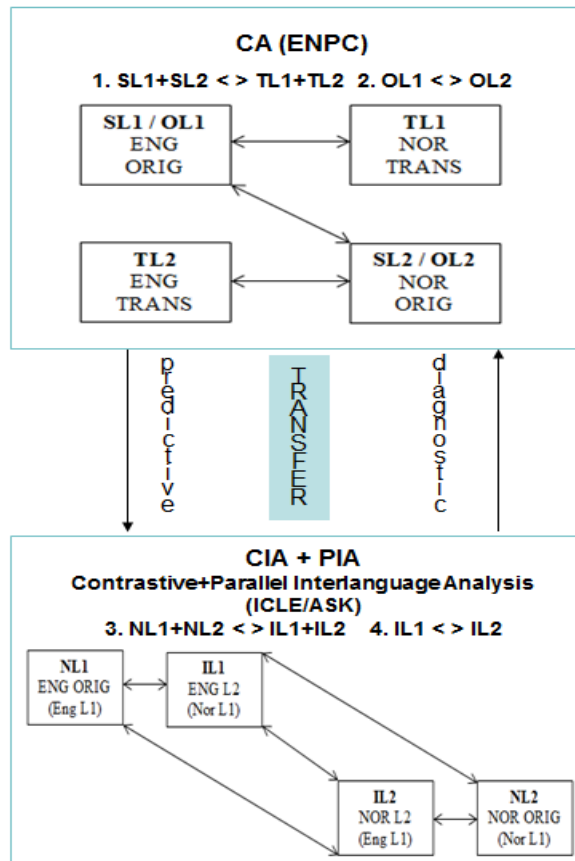
Figure 3 The Parallel Contrastive Model

The "new" model involves four main steps:

1. Contrastive Analysis based on bidirectional translation data (SL1+SL2 ⇔ TL1+TL2)
2. Contrastive Analysis based on comparable original (L1) data (OL1 vs. OL2)
3. Contrastive Interlanguage Analysis based on comparable NS and learner data (NL1+NL2 vs. IL1+NL2)
4. Parallel Interlanguage Analysis based on comparable learner data from two languages (IL1 vs. IL2)

The cross-linguistic comparison performed in steps 1 and 2 circumvents two of the main disadvantages often associated with parallel (translation) and comparable corpora, viz. limited availability of texts (translated language) and text type comparability (OL1-OL2 equivalence) (Granger, 2010).

The first step ensures that we compare like with like, as a solid *tertium comparationis* (TC) is present. By *tertium comparationis* we understand "some kind of constant serving as the background of sameness against which the differences are to be measured" (Ringbom 1994: 738). Thus, an objective "relationship between a unit in the source language and its translation in the target language" is established (Granger 2010: 5). Although it has been claimed that translations "cannot but give a distorted picture of the language they represent" (Teubert 1996: 247), a combination of parallel (translation) and comparable corpora is endorsed by e.g. Johansson (2007) and Granger (2010), as well as by the present authors. The danger of translation bias is counteracted by fact that the corpus is bidirectional, so that mutual correspondence can be taken into account (Altenberg 1999). Thus, step 2 draws on comparable original material in the two languages under study, ensuring that items occur in authentic and natural contexts. For the purpose of this study, we draw on the English-Norwegian Parallel Corpus for both steps 1 and 2, exploiting the fact that we can identify the translation paradigms (horizontal lines in Figure 3) for further scrutiny in the comparable data (the slant line in Figure 3).

The results of the contrastive analysis performed in steps 1 and 2 form the basis for hypotheses about learner behaviour in both of the languages, to be tested in steps 3 and 4. In principle, according to the (strong version of) the Contrastive Analysis Hypothesis (Wardhaugh 1970), any cross-linguistic differences emerging from the contrastive analysis should cause problems for both learner groups, while similarities will facilitate transfer, mainly of the positive kind (e.g. Ringbom 2007) in both interlanguage varieties. Hence, if a linguistic feature is found to differ between Language X and Language Y, that feature should also differentiate L1 and L2 production in both languages. However, as reported by Ortega (2009: 32), a cross-linguistic difference need not lead to problems in both directions of learning, as in the case of the placement of object pronouns in English and French (*I see them* vs. *Je les vois*), which causes more problems for English-speaking learners of French

than French-speaking learners of English.[5] Thus, while keeping in mind the fact that other factors than cross-linguistic differences may be the source of ease or difficulty of L2-learning,[6] we argue that the proposed model is well suited for studies of potential influence of interlingual factors on L2 output, in a way that may nuance the picture that emerges from studies where only one type of interlanguage is studied.

The fourth step—the cross-linguistic analysis of two different interlanguages—represents the most innovative part of the model. The comparison relies on the *tertium comparationis* developed in the first two steps of the analysis; i.e. the (bidirectional) translation paradigms ensure that the cross-linguistic comparison based on comparable corpora is valid. The default hypothesis for this L2 comparison—provided the learners master a particular feature in a target-like fashion—is that the cross-linguistic similarities or differences reflect those uncovered in the L1 comparison. An additional possibility is that the comparison of different L2s may exhibit general (language-independent) features of interlanguage, in much the same way as the comparison of translations into different languages are assumed to display properties that are specific to translated language (e.g. Johansson & Hofland 1994: 27).

Before we outline the method emerging from the model, it should be mentioned that, while the original Integrated Contrastive Model does not necessarily rule out the comparison of different L2s (interlanguages), we have come across very few studies that do this. Exceptions include Demol & Hadermann (2008) and Vanderbauwhede (2012), which will be discussed in Section 2.2, and of which the latter comes closest to the model outlined in the present paper.

---

[5] It may be noted that a learner's L1 is not the only potential source of cross-linguistic influence: influence may take place between an L2 and an L3, for example. However, it lies outside the scope of the present study to pursue this further (but see e.g. Jessner 2006).

[6] Among other factors that have been proposed are markedness (see Ortega: 2009: 37; Jarvis & Pavlenko 2007: 186), perceived language distance (Ellis 2008: 397) and general level of proficiency (Jarvis & Pavlenko 2007: 201).

2.2 Previous ICM-based research analysing two different interlanguages (L2s)

Demol & Hadermann (2008) explore discourse organisation in Dutch and French in terms of parataxis and hypotaxis and the types of subordinate clauses used in the latter. Their material consists of comparable corpora of L1 and L2 Dutch and French, the Dutch L2 coming from French-speaking learners and vice versa. Based on previous cross-linguistic studies of clause linking, Demol & Haderman expected to find more subordination and longer sentences in L1 French than in L1 Dutch, and transfer from the L1 in the L2 varieties. The results were, however, inconclusive: both learner groups produced on average fewer complex and multiple sentences than the L1 groups, but there were no major cross-linguistic differences in either informant group. What distinguishes the method of this study from ours is that it does not use a bidirectional parallel corpus (2008: 256).

In a paper evaluating the effectiveness of the "Integrated Contrastive Model for describing real language use and predicting correct and incorrect L2 productions", Vanderbauwhede (2012) investigates the use of French and Dutch demonstrative determiners in L1 and L2 language production. In line with the ICM, a contrastive analysis on the basis of comparable and translation data is carried out to arrive at accurate descriptions of the determiner systems in the two L1s compared. The CA proper is followed by a Contrastive Interlanguage Analysis, in which native-speaker (reference variety) Dutch is compared with L2 Dutch and native-speaker (reference variety) French with L2 French, in order to establish the impact of the L1 on L2 determiner use. It is explicitly stated that the CIA part of the study only focuses on a comparison of "native and interlanguage varieties of French and Dutch" (2012: 394), i.e. the NL vs. IL branch of the ICM. Our model resembles that of Vanderbauwhede in that it tackles two L1s and two L2s, in contrast to more traditional interlanguage studies using the ICM which typically focus on one L2.

What sets the model proposed in the current paper apart from Vanderbauwhede's, however, is the additional, direct comparison between two different L2s (in our case L2 English vs. L2 Norwegian). It is also worth mentioning that, with a focus on errors and negative transfer in L2 production, Vanderbauwhede's ultimate aim is pedagogical in nature, with the results from the error analysis feeding directly into the development of pedagogical materials for learners of

Dutch and French. The concern of the present paper, in contrast, is to extend the model of comparison so as to facilitate even more robust interlanguage studies which resemble bidirectional contrastive studies by drawing on parallel learner output in different L2s.

2.3 The corpora

This section introduces the primary data, i.e. the corpora, needed for the Parallel Contrastive Model to be operational. By having a concrete idea of what kind of material is required to fulfil the model's potential, we can avoid the pitfalls that may be associated with outlining an entirely abstract model.

The primary data need to be carefully selected and matched at different levels. First, and as indicated above, the model presupposes the following types of corpora representing the same pair of languages:

- a parallel corpus of bidirectional translation data
- comparable learner corpora of two interlanguage varieties representing different languages

For the purposes of illustrating the model, our bidirectional translation corpus is the fiction part of the English-Norwegian Parallel Corpus (ENPC), which consists of original texts in English and Norwegian with translations into the other language (Johansson 2007: 10–15). The original and translated texts are written by professional writers and translators, respectively. The ENPC is used for both for the initial contrastive analysis of L1 data and as a reference variety (L1) corpus (cf. Granger 2015). In order for the two learner corpora to match the ENPC in terms of languages, we need a corpus of English interlanguage produced by Norwegian-speaking learners and a corpus of Norwegian interlanguage produced by English-speaking learners. The Norwegian component of the International Corpus of Learner English (ICLE-NO) and the Anglophone component of Norsk andrespråkskorpus ("Norwegian second language corpus", ASK-ENG)[7] were chosen to

---

[7] Searches were made in texts by learners from the following English-speaking countries: Australia, Canada, Ireland, New Zealand, UK, and USA.

represent the two different interlanguages.[8] In terms of contrastive analysis, the learner corpora are regarded as *comparable corpora* (Johansson 2007: 9), matched by the variables text type (argumentative) and second-language writing.

We must admit that the selected corpora are less than ideally matched, for example with regard to text size, text type (for the L1-L2 dimension) and proficiency level (in the case of the learner corpora). However, no existing set of corpora are perfectly matched in these respects, so for the sake of illustrating the model we decided to proceed with what was available to us, though mindful of the fact that the linguistic features examined may be affected by register. Thus, the Parallel Contrastive Model will be illustrated by a study that compares fiction (ENPC) with argumentative writing (ICLE-NO/ASK-ENG) in corpora of different sizes.[9]

ENPCfiction contains 1.6 million words altogether (divided across four sub-corpora of English originals, Norwegian originals, English translations and Norwegian translations, each made up of 30 text extracts of approx. 12,000 words), whereas ICLE-NO contains around 212,000 words (316 argumentative texts of varying lengths, typically around 600 words each) and ASK-ENG around 75,400 words (175 argumentative texts of varying lengths, typically around 400–500 words each). With regard to L2 proficiency, the learners in ICLE-NO are generally at a higher level (B2/C1/C2, see Granger et al. 2009: 12) than those in ASK-ENG (B1/B2). In an attempt to counter these differences to some extent, we have identified a lexical item for the illustrative study that is general and widespread in both languages and text types as well as across proficiency levels in the learner corpora (see Section 3).

---

[8] For further information on the corpora, see the URLs given at the end of this paper.

[9] The ENPC contains non-fiction, too. However, this part of the corpus is very diverse, spanning from biography to legal texts, but notably lacking argumentative texts. Hence, the more homogeneous fiction part was preferred in the present study.

3. Illustration of method: A Parallel Contrastive Analysis of PEOPLE-*nouns in English and Norwegian L1 and L2*

To illustrate the Parallel Contrastive Model we use the general noun *people* and its Norwegian correspondences to study lexical choice and noun modification patterns in English and Norwegian L1 and L2. We emphasize (i) frequency and types of modifiers of these nouns in the English-Norwegian Parallel Corpus; (ii) the extent to which the patterns gleaned from the parallel corpus can be recognized in English L2 by Norwegian learners and Norwegian L2 by learners whose L1 is English, and (iii) whether the learners have anything in common across languages.

As indicated above, the available corpora are fairly small and somewhat unmatched. It was therefore essential to identify a relatively general object of study that is also frequent in the two languages compared. The noun *people* was selected on the basis of a bottom-up approach to identifying a frequent English noun which often occurs with modification in both the ENPC and ICLE-NO, and which is widely distributed across the texts. Noun modification was deemed suitable for the investigation, not only because of its general frequency, but also because NPs have been described as "a yardstick of syntactic complexity across the variables of development, genre, modality, and cross-linguistic variation" (Ravid & Berman 2010: 6). A challenge in the current context is that NP complexity is also known to vary across registers (e.g. Staples et al. 2016.). However, data from the British National Corpus show that *people*, being one of the most frequently modified nouns in the corpus, occurs with both pre- and postmodification across all the text types.

With the lexeme *people* as our starting point, we first turn to the ENPC to establish the translation paradigm of *people*, i.e. identify Norwegian words to compare *people* with. Then we exploit the bidirectionality of the corpus by searching for English correspondences of the most frequent Norwegian counterparts of *people*. This is followed by a manual analysis of the modification patterns of the selected nouns in the comparable original texts in the ENPC. On the basis of this cross-linguistic comparison, hypotheses for learner behaviour in both languages can be formulated.

The Contrastive Interlanguage Analysis (CIA) starts with the extraction and manual analysis of the same nouns, including their modification patterns, in the learner corpora (ICLE-NO and ASK-ENG),

and is followed by a comparison with the respective L1 data as well as the respective mother tongues of the learners. Finally, in the Parallel Interlanguage Analysis (PIA), or the IL1 vs. IL2 comparison, we can explore similarities and differences between learner English and learner Norwegian.

This final step is in many ways an experiment, as it is not altogether clear to what extent PIA is a viable extension of the Integrated Contrastive Model. However, the intended purpose of the Parallel Interlanguage Analysis is to enable us to explore whether the same similarities and differences can be attested in comparable L2 varieties as in comparable L1 varieties.[10] Thus, the potential and robustness of the Parallel Contrastive Model lies in the fact that the same contrastive method can be applied not only to the Contrastive Analysis proper but also to the Parallel Interlanguage Analysis, i.e. the OL1 vs. OL2 branch of the CA and the IL1 vs. IL2 branch of the PIA in Figure 3. Both are anchored in the belief that translation is a good *tertium comparationis* that can form the basis for traditional comparable contrastive studies (Johansson 2007: 5) as well as for comparable interlanguage analysis.

3.1 Contrastive analysis

We will now turn to the practical aspects of the model, following the various steps outlined in Section 2.1. This illustrative case study starts with the English-Norwegian contrastive analysis of *people* and its modification patterns.

3.1.1 Translation paradigm of people *in the ENPC*

The translation paradigm of *people* in the ENPC is shown in Table 1, where it can be observed that there are two main Norwegian

---

[10] We are aware of two projects involving a similar mix of data, notably translation and interlanguage data; however, they focus mainly on potential similarities between translation and interlanguage production: Halverson et al., *Multikompetanse og tospråkleg modus* 'Multi-competence and bilingual modes' (https://www.hvl.no/en/research/group/alm/) and Behrens et al., *Språk som produkt og prosess* 'Language as product and process' (http://www.hf.uio.no/ilos/english/research/projects/language-as-product-and-process/).

correspondence types, namely *mennesker/menneskene* (i.e. plural indefinite/definite "bokmål" forms of the noun MENNESKE)[11] and the lemma FOLK. Together they account for almost 75% of the correspondences (353 occurrences out of a total of 484).[12] The most frequently occurring nouns other than *mennesker/menneskene* and FOLK were *personer* 'persons' (8) and compounds including *folk* (8), e.g. *fattigfolk* Lit: 'poorpeople'.

Table 1 Translation paradigm of *people* (ENPC)

Norwegian translations of *people*	Raw #	
mennesker/menneskene	153	**353**
FOLK	200	
Noun other than *menneske**/FOLK[13]	34	
Other	76	
Ø	21	
TOTAL	484	

With a bidirectional translation corpus such as the ENPC, we can also start in the Norwegian original texts to discover to what extent *people* is the translation of *menneske** and FOLK. Indeed, Table 2 reveals that in 375 out of 517 instances (72.5%), *people* is used to translate these two nouns.[14] The most recurrent nouns other than *people* were *folk/folks* (13) and *man/men* (13). The single-most prominent 'other' translation correspondence was the indefinite pronoun *everybody*/*everyone* with 10 occurrences.

---

[11] Norwegian has two written standards, "bokmål" and "nynorsk", which may differ slightly in orthography and morphology, but are mutually intelligible); see e.g. Vikør (2015). All the Norwegian translations are written in "bokmål". Two of the Norwegian original texts are written in "nynorsk"; however, we have chosen to include only the "bokmål" forms of plural *menneske* in this study. The main reason for this is that the plural indefinite form in "nynorsk" coincides with the singular indefinite form *menneske*.

[12] In this part of the study we refer only to overall frequencies. Importantly, however, the instances of *people* are distributed across all the 30 original texts.

[13] Only the two plural forms (*mennesker/menneskene*) of the lemma MENNESKE are represented in the correspondences of *people*; *menneske** is used as a shorthand to indicate this.

[14] The raw frequencies reported in Tables 1 and 2 are more or less comparable, as the sub-corpora of the ENPC contain roughly the same number of tokens.

Table 2 Translation paradigm of *menneske*/*FOLK[15]

English translations of *menneske*/*FOLK	Raw #
people	375
Noun other than *people*	77
Other	32
Ø	33
TOTAL	517

In the following examples, the main correspondence trends are shown in (1)–(2), in which *people* translates to *menneske*/*FOLK and vice versa, while examples (3) and (4) show other correspondence types. In (3) the combination *old people* has been translated into a nominalized adjective, (*de*) *gamle* '(the) old', while in (4), *menneskene* has been rendered by the personal pronoun *they*.

   (1)   I can't be expected to do the work of two *people*. (AB1)[16]
         De kan ikke vente at jeg skal gjøre jobben for to *mennesker*.
         (AB1T)

   (2)   Det er merkelig hvor likt *folk* tenker … (KF1)
         It's strange how similarly *people* think … (KF1T)

   (3)   The pride the *old people* took in him was not just the snobbery of
         the poor and uneducated, … (NG1)
         Den stoltheten de *gamle* følte over ham var ikke bare den vanlige
         snobbete gleden hos de fattige og uopplyste … (NG1T)

   (4)   *Menneskene* måtte selv ta del i kampen mot det onde. (JG1)
         They had to act for themselves in the struggle against evil.
         (JG1T)

The overviews given in Tables 1 and 2 make it possible to calculate the intertranslatability, or Mutual Correspondence (Altenberg 1999), of the

---

[15] *Menneske** is distributed across all the 30 original texts, while FOLK is found in 28 texts.
[16] The ENPC corpus ID identifies the author by initials (AB = Anita Brookner) and the text by that author (1). Translations are marked with a T. For an overview of the texts in the ENPC, see Johansson (2007: 329–334).

three nouns under investigation. Mutual Correspondence (MC) is a simple measure of how often items are translated by each other in a bidirectional translation corpus, and is calculated by adding the number of times each item is translated by the other, divided by the total number of occurrences of the items in the corpus and multiplied by 100 to give a percentage. The Mutual Correspondence of *people* and *mennesker*/*menneskene*/FOLK is as high as 72.7%.

As this study is mainly concerned with instances of the PEOPLE-nouns when they are modified, the MC of the modified instances of the nouns was also calculated, showing an almost equally high MC score of 68.9%. We take this MC as a good indication that these nouns, with and without modification, are objectively comparable items in English and Norwegian.

3.1.2 Modification patterns of the nouns (people, menneske*, *FOLK*)
The overall proportion of modification of *people* in the English original texts and *menneske** and FOLK in the Norwegian original texts is shown in Figure 5. In the case of *people* and *menneske** roughly 50% are modified, while FOLK is modified in around 30% of the cases. In the English material, all the 30 texts in the corpus use *people* with modification, albeit with varying frequency, ranging from one occurrence to 33 per text. In the Norwegian data, *menneske** is used with modification in 25 out of the 30 texts (0-15 occ. per text), while FOLK is used with modification in 26 texts (0-9 occ. per text).
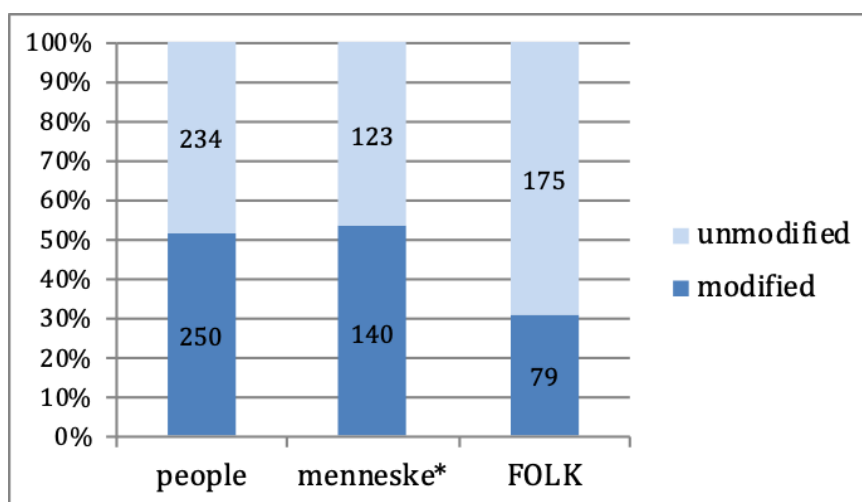
Figure 5 Proportion (and raw numbers) of modified vs. unmodified
people/*menneske**/FOLK in the original texts in the ENPC

The data are further broken down into type of modification and Table 3 shows the overall distribution of pre- and postmodification of the nouns, as well as the combination of the two modification types.[17]

Table 3 General modification patterns of the PEOPLE-nouns in the ENPC

Type of modification	*people*	*menneske**	FOLK	*menneske**+ FOLK
Premod.	80 (32%)	59 (42.1%)	17 (21.5%)	76 (34.7%)
Postmod.	138 (55.2%)	61 (43.6%)	59 (74.7%)	120 (54.8%)
Pre+postmod.	32 (12.8%)	20 (14.3%)	3 (3.8%)	23 (10.5%)
Total	250	140	79	219

From Table 3 it is interesting to note that *menneske** and FOLK differ not only in the frequency with which they are modified overall (see Figure 5), but also in how often they are used with pre- and postmodification,

---

[17] As pointed out by one of the anonymous reviewers, *people* is polysemous in the sense that it has an uncountable and countable use, as well as a specific and generic use. FOLK has a similar profile. Although this may have an impact on the distribution of modifiers, all uses, with the expcetion of countable *people*/ FOLK (e.g. *the relics of a people*), are represented with modification in the material studied here.

respectively. English *people* is somewhere in between and close to the Norwegian nouns if these are merged (right-most column in Table 3).

As this study is merely meant to illustrate a model, we have chosen to delimit the remainder of it to the most frequent type of modification. Thus, the dataset will consist of those noun phrases which have postmodification only; in this way the interference of a premodifier as a potential variable for postmodifier use will also be avoided.

In terms of distribution and dispersion, it is important to note that *people* is used with postmodification in 28 texts, and the number of occurrences per text ranges from 1 to 16. The Norwegian texts mirror the distribution and dispersion in the English material, with 29 texts using postmodified *menneske** or FOLK, ranging from 1 to 14 occurrences per text. Both populations contain one outlier each (with 16 and 14 occurrences, respectively); if these are removed from the postmodification row in Table 3, *people* is postmodified in 52% (instead of 55.2%) of the cases, while *menneske**/FOLK combined are postmodified in 51.7% (instead of 54.8%). Table 4 displays these updated figures, i.e. 122 cases of postmodified *people*, and 106 cases of postmodified *menneske**/FOLK (instead of the 138 and 120 shown in Table 3).

Two postmodification patterns stand out as the most frequent options in both languages, namely relative clause and prepositional phrase (PP), exemplified in (5) and (6), respectively. These are the two patterns we will focus on.

(5) Gjestens blikk flakket og forsøkte å finne feste blant *menneskene som satt ved bordene omkring*. (FC1)
The guest's eye wandered and tried to find a mooring among the *people who sat at the tables around them*. (FC1T)

(6) *People from Toronto* used to go to Buffalo for the weekends, ...
(MA1)
Folk fra Toronto pleide å dra over til Buffalo i helgene, …
(MA1T)

Proportionally, and as shown in Table 4, *people* is less frequently modified in either of these ways compared to the two Norwegian nouns put together. However, they are still clearly the main postmodifier

options for *people*. Again it can be observed that *menneske** and FOLK behave slightly differently in their preferred modification types: relative clause postmodification is more frequent with *menneske** and prepositional phrases with FOLK. However, these numbers are too low to draw any general conclusions regarding differences in modification patterns between *menneske** and FOLK.

Table 4 Postmodification patterns the PEOPLE-nouns in the ENPC (original texts) (without outliers)

Postmodifier	*people*	*menneske**	FOLK	*menneske**+FOLK
Relative clause	47 (38.5%)	28	20	48 (45.3%)
PP	45 (36.9%)	24	30	54 (50.9%)
Other	30 (24.6%)		4	4 (3.8%)
Total (without outliers)	122			106

The 'other' category represents a more varied set of postmodification types for English than Norwegian, mainly due to two types of non-finite clause modification which are more or less ruled out in Norwegian for syntactic reasons, i.e. present and past participle clauses, as shown in examples (7) and (8). Of the four postmodifiers in the 'other' category in the Norwegian material, three are instances of infinitive clauses, e.g. (9), and one is a very rare example of a present participle clause, (10).

(7)   He'd see to it that all the people *working in the outer rooms* got cakes and drink. (JC1)

(8)   I simply cultivate relationships with people *plugged into the system at various points*. (SG1)

(9)   Even hadde hørt folk *fortelle det*; og nå hørte han og så det selv. (KAL1)
        Espen had heard people *talk about it*. Now he had heard it and seen it himself. (KAL1T)

(10) JEG SER DEM ALLTID SOM FOLK *VENTENDE REGN*
      MOSE BARN (CL1)[18]
      I ALWAYS SEE THEM AS PEOPLE *WAITING FOR RAIN*
      MOSS CHILDREN (CL1T)

3.2 From CA to PIA

The contrastive analysis performed above has uncovered a lot of similarity in the uses of English *people* and Norwegian *menneske*/*FOLK, including cases in which the nouns are postmodified. These observations should be encouraging for learners of both languages, in the sense that there is potential for positive transfer from their mother tongue. It is, however, important to note that the CA also indicates that the two learner groups are, to a certain degree, faced with different challenges:

- Norwegian learners of English may fail to use postmodifiers other than relative clauses and PPs due to the shortage of these in Norwegian L1.
- English-speaking learners of Norwegian may transfer participle-clause modifiers from their L1 (at the cost of relative clauses and PPs).

In addition, we do not know whether the learners will (subconsciously) equate *people* with *menneske** or FOLK or both. As shown in Table 3, FOLK has proportionally less premodification and proportionally more postmodification than *menneske**, and the two Norwegian nouns show slight differences with regard to the type of postmodification they prefer (Table 4). When the focus is now shifted from Contrastive Analysis, via Contrastive Interlanguage Analysis, to Parallel Interlanguage Analysis, it will be interesting to see how the two learner groups tackle these potential challenges.

3.3 Analysis of the interlanguage varieties

This section presents an analysis of the L2 varieties following the same structure as that of the L1 varieties given above. To avoid repetitiveness,

---

[18] Capital letters used in the original text.

the Contrastive Interlanguage Analysis (comparison with L1) and the Parallel Interlanguage Analysis (L2 English vs. L2 Norwegian) are presented side by side. The Figures and Tables in Section 3.3.2 show L2 patterns only, with reference to Section 3.1.2 for comparison with L1.

3.3.1 Hypotheses for the Contrastive and Parallel Interlanguage Analysis

As pointed out in Section 3.2, the similarity revealed in modification patterns of PEOPLE-nouns in Norwegian and English leads us to believe that the learners will not have great trouble with these nouns. That is, learners of both languages can benefit from the cross-linguistic similarities between English and Norwegian which facilitate positive transfer (Ringbom 2007). The differences uncovered are to a large extent connected to different frequencies of similar constructions. However, the use of postmodifier types other than relative clauses and prepositional phrases reflects a syntactic difference between the languages, namely the presence of adnominal participle clauses in English, but (typically) not in Norwegian. This gives rise to the following hypotheses for the contrastive and parallel interlanguage analyses:

1. Postmodifying relative clauses and PPs will be overrepresented in L2 English (compared to L1 English).
2. Postmodifying relative clauses and PPs will be underrepresented in L2 Norwegian (compared to L1 Norwegian).
3. Postmodifier types other than relative clauses and PPs will be used more in L2 Norwegian and less in L2 English than in the respective L1s.
4. The potential differences in the modification patterns of FOLK and *menneske** will not be reflected in L2 Norwegian.

To some extent the cross-linguistic differences discovered in the CA should have opposite effects on the two interlanguages. This is the basis for the first three hypotheses. Because Norwegian L1 uses more relative clauses and PPs than English we expect that these modifier types will be overrepresented in L2 English and underrepresented in L2 Norwegian.

The fourth hypothesis concerns L2 Norwegian only since the two Norwegian PEOPLE-nouns have somewhat different modification patterns

(see Tables 3 and 4). However, as an effect of this difference, the learners of English may base their use of *people* on the patterns of FOLK, *menneske** or a mix of the two.

3.3.2 Occurrence and modification of PEOPLE-nouns in learner and native language

People, menneske* and FOLK are all widespread in the learner corpora: *people* occurs in 88% of the ICLE-NO texts (n = 1340), and *menneske** and/or FOLK in 80% of the ASK-ENG texts (n = 184+226).[19] The proportion of modification in the learner texts is shown in Figure 8.
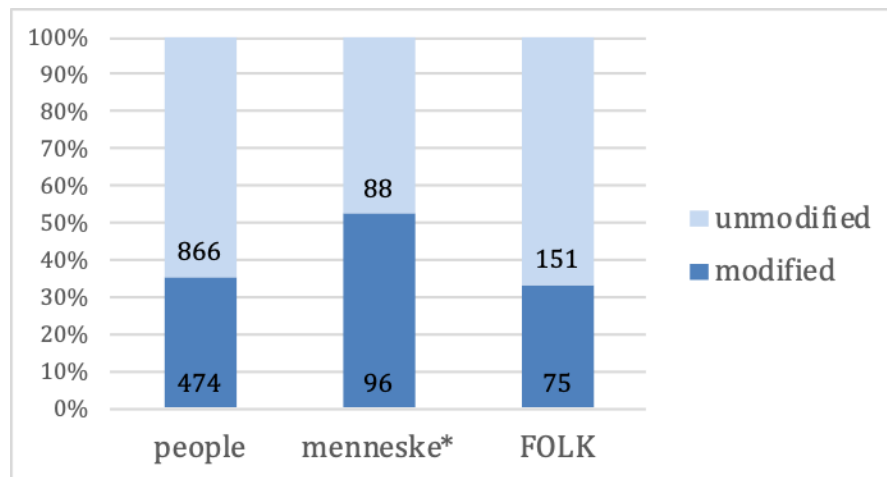

Figure 8 Proportions (and raw numbers) of modified vs. unmodified *people*, *menneske** and FOLK in L2 English and Norwegian (ICLE-NO and ASK-ENG)

For *people* and FOLK, the majority of instances are unmodified, while *menneske** is modified just over 50% of the time. At this general level, the learners represented in ASK-ENG thus appear to have grasped the different modification patterns of *menneske** and FOLK apparent in Figure 5 above. The proportion of modified *people* in ICLE-NO is lower than that of L1 *people* in the ENPC, and is in fact rather similar to L1 Norwegian FOLK. It is thus possible that the learners of English equate

---

[19] The texts in ASK-ENG are all in "bokmål".

people with FOLK rather than with *menneske**. The learners of Norwegian, on the other hand, may seem to equate *people* with *menneske*,* as far as the use of modification is concerned.

When a distinction is made between modifier types, a more nuanced picture emerges, as shown in Table 5, which is comparable to Table 3 above. We may note that the most complex pattern with both pre- and postmodification, as in (11), is less frequent than in L1 in both learner corpora, most noticeably in L2 Norwegian.[20]

> (11)   We need an army with <u>*sensible, young* people <u>with a wish to restore peace</u></u>…   (ICLE)

Table 5 General modification patterns of the PEOPLE-nouns in L2 English and Norwegian

	ICLE-NO	ASK-ENG		
Type of modification	*people*	*menneske**	FOLK	*menneske** + FOLK
Premod.	161 (34.0%)	52 (54.2%)	29 (38.7%)	81 (47.4%)
Postmod.	276 (58.2%)	39 (40.6%)	46 (61.3%)	85 (49.7%)
Pre+postmod.	37 (7.8%)	5 (5.2%)	0	5 (2.9%)
Total	474	96	75	171

The proportions of modification types in ICLE-NO are similar to those in L1 English in the ENPC, and thus also to the combined figures for *menneske** + FOLK in L1 Norwegian. These similarities may of course be due to either a grasp of native-like modification patterns for *people* or to mixed transfer from the patterns of *menneske** and FOLK.

L2 Norwegian has higher proportions of premodifiers than L1 Norwegian as well as L1 and L2 English, and by implication lower proportions of postmodifiers. However, this is mainly due to the pattern of *menneske**: the proportions of modifier types are fairly similar between *people* and FOLK. More specifically, the learners in ASK-ENG use premodification more often with *menneske** than with FOLK. This is also the case in L1 Norwegian in the ENPC (Table 3), but the proportion of premodification is higher in L2 than in L1 with both nouns. Thus, the

---

[20] Examples from the learner corpora have been rendered verbatim with any infelicities intact.

apparently native-like proportions of modified and unmodified FOLK/*menneske** needs to be nuanced.

We will now take a closer look at the postmodifier types used in the learner corpora. As in section 3.1.2, the postmodifiers that occur in conjunction with a premodifier have been excluded from this part of the study. Because the texts in ICLE-NO and ASK-ENG are rather short (Section 2.3), it is unremarkable that not all of them contain a postmodified PEOPLE-noun: *People* occurs with postmodification in 158 ICLE texts (50%) and *menneske**/ FOLK + postmodifier in 62 ASK texts (35.4%). Frequencies per text vary from 1 to 10 occurrences in ICLE and from 1 to 5 in ASK, but few texts in either corpus contain more than one or two.[21] To ensure comparability of the ICLE and ASK data, we decided to use one (random) occurrence per text for the cross-linguistic comparison of postmodifier choice. The resulting reduced dataset consists of 158 and 62 instances of postmodified PEOPLE-nouns.

Table 6 Postmodification patterns in L2 English and Norwegian (represented by a single occurrence per text containing a postmodified PEOPLE-noun)

	ICLE-NO	ASK-ENG		
Postmodifier	*people*	*menneske**	FOLK	*menneske**/FOLK
Relative clause	65 (41.1%)	18	25	43 (69.4%)
PP	59 (37.3%)	7	6	13 (21.0%)
Other	34 (21.5%)	2	4	6 (9.7%)
Total	158	27	35	62

Table 6 presents the postmodifier types found in the reduced L2 datasets. It should be noted first of all that the learners of both languages use postmodifiers only from the repertoire that is available in the target language. That is, we have not found any postmodifier types that are ungrammatical. This no doubt reflects the similarities in syntactic potential between the languages, but may also be due to perceptions of cross-linguistic differences: for example, there was no evidence of English-speaking learners of Norwegian transferring postmodifying *-ing* clauses, contrary to our third hypothesis (Section 3.3.1). Any differences

---

[21] The cut-off point for outliers is 3 in ICLE-NO and 1 in ASK-NO, i.e. texts with more than 3 and 1 occurrences of a postmodified PEOPLE-noun, respectively.

between L1 and L2 patterns are thus related to diverging selections of modification types, not to error.

In both interlanguage varieties, relative clauses are the most frequent choice. This is not unexpected for learner English, since relative clauses are also a frequent choice in L1 Norwegian. The proportion of relative clauses in L2 English lies in between that of L1 Norwegian and L1 English (and is thus similar to both). L2 Norwegian displays a less expected pattern, with relative clauses being the most frequent postmodifier type, by a much wider margin than in both L2 English and L1 Norwegian. Examples of relative clause postmodifiers from the learner data are given in (12) and (13).[22]

> (12)   … folk *som har "gode" jobber og tjener penger er folk som vi tror lykkes og har suksess*, … (ASK-ENG) [people who have good jobs and earn money are people who we think succeed and have success]

> (13)   Bring in *people that know how the system really works*. (ICLE-NO)

A partial explanation for the more frequent use of relative clauses in L2 Norwegian might lie in the fact that Norwegian lacks an equivalent of *-ing* participle clauses, which are the most frequent of the English modifier types included in 'other' in Tables 4 and 6. It may thus be natural for an English-speaking learner of Norwegian to use a finite relative clause in contexts where a postmodifying *-ing* clause would have been used in English.

The second most frequent postmodifiers in both corpora are prepositional phrases, as exemplified in (14) and (15).

> (14)   Det er veldig hyggelig hvis *menneskene i et sted* er vennlig og imøtekommende. (ASK-NO) [It is very nice if *people in a place* are friendly and welcoming]

---

[22] Examples in Norwegian are followed by a fairly literal translation given in square brackets.

(15)   In a way I feel sorry for *people* <u>*with no dreams*</u>. (ICLE-NO)

It was hypothesized in Section 3.3.1 that prepositional phrases would be overrepresented in L2 English and underrepresented in L2 Norwegian due to their respective frequencies in the learners' first languages (Section 3.1.2). The material analysed here did not support the hypothesis about L2 English; prepositional phrases and relative clauses in ICLE-NO have fairly similar frequencies, and the pattern resembles that of English L1 as well. By contrast, the frequency of PPs in learner Norwegian is even lower than expected, compared to both L1 English and L1 Norwegian. It seems that the learners of Norwegian select a relative clause for postmodification almost by default, thereby overgeneralizing this realization type to most instances of postmodification.[23]

Comparing Tables 4 and 6, we see that the postmodification patterns of *people* are very similar across L1 and L2, while the patterns of *menneske** plus FOLK are more different as to the proportions of relative clauses and PPs. This is mostly due to FOLK, which prefers PP postmodifiers in L1 and relative clauses in L2. It needs to be further investigated whether results of the L1/L2 comparison might be partly due to the fact that the L1 texts come from fiction and the L2 texts from argumentative writing. However, the similarities between the patterns found in L1 and L2 English may suggest that register has not played a major role.

In both interlanguage varieties, the postmodifying prepositional phrases are often spatial or denote an attribute of the head noun, as shown in (14) and (15). No particular pattern can be detected with relative clauses. Furthermore, there seem to be no frequently recurring postmodifiers of either type, possibly excepting *people in general* and *people from* + place name */* FOLK/*menneske* fra* + place name.

---

[23] An alternative (or complementary) explanation could be that PPs are more compressed and less explicit noun modifiers than relative clauses (cf. Biber and Gray 2010) and as such may represent a more advanced level of proficiency, as well as a more academic-like register (ibid.).

3.3.3 Summary of findings of the interlanguage analysis

Summing up the findings of the bidirectional CIA, we can conclude that the expected overrepresentation of relative clauses and PPs in English L2 did not occur. Nor did the potential avoidance of other postmodifiers. That is, the choice of postmodifiers of *people* in ICLE-NO is quite similar to the L1 patterns in the ENPC. Compared to L1 Norwegian, the general modification patterns of *people* in ICLE-NO resemble those of FOLK more than those of *menneske**. As for Norwegian L2 (ASK-ENG), the expected underrepresentation of relative clauses was not attested, but the underrepresentation of PPs was even more pronounced than expected. There was no evidence of negative transfer of other postmodifier types. FOLK and *menneske** have different modification patterns in both L1 and L2 Norwegian, with more unmodified FOLK, but less frequent postmodification of *menneske** (among those nouns that occur with modification). However, L1 and L2 users have different preferred types of postmodifiers, with relative clauses appearing to be more of a default option in L2.

3.4 Findings of the parallel contrastive study

To illustrate the Parallel Contrastive Model we focused on a frequent English noun—*people*—and its Norwegian correspondences as evidenced through a bidirectional analysis based on the ENPC. Since *people* corresponds regularly with the Norwegian nouns FOLK and *menneske** in translations between the languages, we can safely compare these nouns across comparable L2 corpora. Furthermore, the study of original texts in the ENPC showed that the most frequently chosen modifiers are fairly similar in English and Norwegian, thus creating a sound basis for the cross-linguistic study of modification patterns of PEOPLE-nouns in ICLE-NO and ASK-ENG. While the contrastive analysis based on bidirectional (comparable and translated) data showed a high degree of cross-linguistic similarity, the differences were considered likely to pose some challenges for both English-speaking learners of Norwegian and Norwegian-speaking learners of English. As detailed above, not all the hypotheses were supported. But although the L2 patterns were not exactly as predicted, the divergences from L1 patterns may still be attributable to the cross-linguistic differences, albeit in different ways than foreseen. For example, the over-reliance on

relative clause postmodifiers in L2 Norwegian might reflect a perception that relative clauses have more uses in Norwegian than in English, resulting in an underrepresentation of PPs. However, this point obviously needs further study.

The Parallel Interlanguage Analysis aimed to uncover the cross-linguistic differences between interlanguage varieties of two languages and furthermore to compare them to the differences between the corresponding L1 varieties. As regards modified vs. unmodified PEOPLE-nouns, the L1 differences do not match the L2 differences: in L1 (ENPC) there are fewer unmodified *people* than FOLK + *menneske**, while in L2 (ICLE and ASK) there are more unmodified *people* than FOLK + *menneske**. The preferred types of postmodifiers are similar in that relative clauses are more frequent in Norwegian than in English in both L1 and L2 varieties. However, the use of PPs shows conflicting cross-linguistic patterns across L1 and L2 varieties: PPs are *less frequent* in L1 English than in L1 Norwegian, but *more frequent* in L2 English than in L2 Norwegian. In brief, the comparison of both L1 and L2 varieties of English and Norwegian confirms the similar syntactic potential in these languages, but also displays how they differ in their preferred ways of saying things. The cross-linguistic differences and similarities are summarized in Table 7.

Table 7 Cross-linguistic differences and similarities between English and Norwegian L1 and L2

	L1 varieties	L2 varieties
Modified vs. unmodified PEOPLE-nouns	More modification in English	More modification in Norwegian
Premodification vs. postmodification	Postmodification is more common in both languages (similar proportions)	More postmodification in English. Similar proportions of pre- and postmodification in Norwegian.
Preferred type of postmodifier	Relative clauses and PPs are evenly distributed in both languages; more 'other' types in English.	Relative clauses much more common and PPs much less common in Norwegian than in English.

4. Concluding remarks

This paper has presented a parallel contrastive model for the cross-linguistic investigation of L1 and L2 varieties of a language pair. The Parallel Contrastive Model (PCM) is not entirely novel: it can be seen as an extension of the practices, albeit not the principles, of the Integrated Contrastive Model (Granger 1996); most previous ICM studies have been concerned with only *one* interlanguage variety. The distinguishing feature of the PCM as compared to the ICM is thus that it explicitly includes two different interlanguage varieties. Hence we can use it to explore the extent to which a contrastive analysis of bidirectional data can provide viable hypotheses for interlanguage production in both languages involved. In addition, the model enables explorations of the extent to which transfer may be asymmetrical (Ortega 2009: 32), i.e. the extent to which cross-linguistic (L1) similarities and differences produce the same amount of ease/difficulty in both directions of language learning. Finally the comparison of L2 varieties of different languages might be able to shed some light on linguistic features that are specific to learners across languages.

The model also extends the practices of the bidirectional parallel corpus model (Johansson 2007) in seeing interlanguage (in addition to translation) in a bidirectional perspective. Crucially, the bidirectional data provide a *tertium comparationis* for the parallel interlanguage analysis. Without this we could not be confident that we are comparing equivalent items in the cross-linguistic L2 analysis based on comparable corpora. The model thus facilitates systematic cross-linguistic comparison of both L1 and interlanguage varieties.

As mentioned above, we must acknowledge that the case study involves some problems of corpus comparability, the most important of which are differences in text types between the L1 and the L2 corpora (fiction / argumentative) and differences in proficiency level between ASK-ENG and ICLE-NO (see Section 2.3). These differences are not trivial as both text type and proficiency level have been shown to influence noun phrase complexity (e.g. Biber & Conrad 2009; Vyatkina 2013). It will thus be advisable for further uses of the model to take these variables into account. However, since the present case study was intended as an illustration of the Parallel Contrastive Model rather than as a full-fledged study of noun phrase complexity in L1 and L2, we felt justified in using the corpora that were available to us.

In spite of these limitations, we believe the case study has illustrated the potential of the Parallel Contrastive Model. The fact that not all the hypotheses based on the initial contrastive analysis were supported does not undermine this; rather it highlights the fact that the relationship between a learner's L1 and the language to be learnt is a complex matter and only one aspect of the process of language learning (Jarvis & Pavlenko 2007: 174 ff). The model draws on the strengths of both the bidirectional model of contrastive analysis and the Integrated Contrastive Model: The bidirectional corpus provides a firm empirical basis for the cross-linguistic analysis, and the *tertium comparationis* that stems from it gives a sound starting point for the L1/L2 comparison. Furthermore, the availability of learner corpora of both languages included in the CA, with the learners' L1 and L2 corresponding to the source and target languages of the translation corpus presents researchers with a unique opportunity to study transferability (and actual transfer) of cross-linguistic similarities and differences in both of the interlanguage varieties concerned. Finally, it enables the study of similarities between translation and L2 production, an avenue which has not been explored in the present paper (but which is alluded to e.g. by Altenberg 2002).[24] Hopefully, then, given the right combination of translation and learner corpora, the Parallel Contrastive Model can provide a new perspective on the interfaces between contrastive analysis, interlanguage and Learner Corpus Research.

---

[24] See also ongoing work by Gujord & Halverson, e.g. https://www.hf.uio.no/multiling/english/news-and-events/events/guest-lectures-seminars/2016/160531_gujordm.halverson.html.

References

Altenberg, Bengt. 1999. "Adverbial connectors in English and Swedish: Semantic and lexical correspondences." *Out of Corpora: Studies in Honour of Stig Johansson*. Eds. Hilde Hasselgård and Signe Oksefjell. Amsterdam: Rodopi. 249–268

Altenberg, Bengt. 2002. "Causative constructions in English and Swedish: A corpus-based contrastive study." *Lexis in Contrast. Corpus-based Approaches*. Eds. Bengt Altenberg and Sylviane Granger. Amsterdam: John Benjamins. 97–118.

Biber, Douglas and Susan Conrad. 2009. *Genre, Register and Style*. Cambridge: Cambridge University Press.

Biber, Douglas and Bethany Gray. 2010. "Challenging stereotypes about academic writing: Complexity, elaboration, explicitness." *Journal of English for Academic Purposes* 9, 2–20.

Demol, Annemie and Pascale Hadermann. 2008. "An exploratory study of discourse organisation in French L1, Dutch L1, French L2 and Dutch L2 written narratives." *Linking up Contrastive and Learner Corpus Research*. Eds. Gaëtanelle Gilquin, Szilvia Papp and María Belén Díez-Bedmar. Amsterdam/Atlanta: Rodopi. 255–282.

Ellis, Rod. 2008. *The Study of Second Language Acquisition*. 2nd ed. Oxford: Oxford University Press.

Gilquin, Gaëtanelle. 2000/2001. "The Integrated Contrastive Model: Spicing up your data." *Languages in Contrast* 3(1) (printed 2003), 95–124.

Granger, Sylviane. 1996. "From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora." *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies, Lund 4-5 March 1994. Lund Studies in English 88*. Eds. Karin Aijmer, Bengt Altenberg and Mats Johansson. Lund: Lund University Press. 37–51.

Granger, Sylviane. 2010. "Comparable and translation corpora in cross-linguistic research: Design, analysis and applications." *Journal of Shanghai Jiaotong University*, 2010. [Available at: http://sites-test.uclouvain.be/cecl/archives/Granger_Crosslinguistic_research.pdf accessed 18 May 2018].

Granger, Sylviane. 2015. "Contrastive Interlanguage Analysis. A reappraisal." *International Journal of Learner Corpus Research* 1(1), 7–24.

Granger, Sylviane, Estelle Dagneaux, Fanny Meunier and Magali Paquot. 2009. *International Corpus of Learner English. Version 2.* Louvain: Presses Universitaires de Louvain.

Jarvis, Scott and Aneta Pavlenko. 2007. *Crosslinguistic Influence in Language and Cognition.* London and New York: Routledge.

Jessner, Ulrike. 2006. *Linguistic Awareness in Multilinguals: English as a Third Language.* Edinburgh: Edinburgh University Press.

Johansson, Stig. 2007. *Seeing through Multilingual Corpora. On the Use of Corpora in Contrastive Studies.* Amsterdam/Philadelphia: Benjamins.

Johansson, Stig and Knut Hofland. 1994. "Towards an English-Norwegian Parallel Corpus." *Creating and Using English Language Corpora: Papers from the Fourteenth International Conference on English Language Research on Computerized Corpora, Zurich 1993.* Eds. Udo Fries, Gunnel Tottie and Peter Schneider. Amsterdam: Rodopi. 25–37.

Lado, Robert. 1957. *Linguistics across Cultures. Applied Linguistics for Language Teachers.* Ann Arbor: The University of Michigan Press.

Ortega, Lourdes. 2009. *Understanding Second Language Acquisition.* London: Hodder Education.

Ravid, Dorit D. and Ruth A. Berman. 2010. "Developing noun phrase complexity at school age: A text-embedded cross-linguistic analysis." *First Language* 30(1): 3–26.

Ringbom, Håkan. 1994. "Contrastive analysis." *Encyclopedia of Language and Linguistics 2.* Eds. Ronald E. Asher and James M.Y. Simpson. Oxford: Pergamon. 737–742.

Ringbom, Håkan. 2007. *Cross-Linguistic Similarity in Foreign Language Learning.* Clevedon: Multilingual Matters.

Staples, Shelley, Jesse Egbert, Douglas Biber and Bethany Gray. 2016. "Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre." *Written Communication* 33(2). 149–183.

Teubert, Wolfgang. 1996. "Comparable or parallel corpora?" *International Journal of Lexicography* 9(3): 238–264.

Vanderbauwhede, Gudrun. 2012. "The Integrated Contrastive Model evaluated: The French and Dutch demonstrative determiner in L1 and L2." *International Journal of Applied Linguistics*, Vol. 22, No. 3, 392–413.

Vikør, Lars. 2015. "Norwegian: Bokmål vs. Nynorsk." Available at http://www.sprakradet.no/Vi-og-vart/Om-oss/English-and-other-languages/English/norwegian-bokmal-vs.-nynorsk/, accessed 18 May 2018].

Vyatkina, Nina. 2013. "Specific syntactic complexity: developmental profiling of individuals based on an annotated learner corpus." *The Modern Language Journal*, Vol. 97: 11–30.

Corpora

British National Corpus (BNC). Davies, Mark. (2004-) *BYU-BNC*. (Based on the British National Corpus from Oxford University Press). Available online at http://corpus.byu.edu/bnc/. [accessed 10 April 2018], see also http://www.natcorp.ox.ac.uk/ [accessed 11 April 2018].

English-Norwegian Parallel Corpus (ENPC): http://www.hf.uio.no/ilos/english/services/omc/enpc/ [accessed 2 May 2018].

International Corpus of Learner English (ICLE): https://uclouvain.be/en/research-institutes/ilc/cecl/icle.html [accessed 2 May 2018].

Norsk andrespråkskorpus (ASK): http://clarino.uib.no/ask/ask [accessed 2 May 2018].