# Contrastive analysis and learner language:

# A corpus-based approach

**Stig Johansson**

University of Oslo
2008

## Preface

The present text has grown out of several years' work on the English-Norwegian Parallel Corpus (ENPC), as presented in my recent monograph *Seeing through Multilingual Corpora: On the use of corpora in contrastive studies* (Benjamins 2007). It differs from the previous book in having been prepared as a textbook for an undergraduate course on 'Contrastive analysis and learner language'. Though there is some overlap between the two texts, the coverage in the textbook is wider. The main difference is that the emphasis is both on contrastive analysis and on the study of learner language, with the aim of showing how these fields are connected. The textbook is intended to be used in combination with a selection of papers on individual topics, some of which go beyond corpus analysis.

I am grateful to my co-workers in the project leading to the ENPC, in particular: Knut Hofland, University of Bergen, and Jarle Ebeling and Signe Oksefjell Ebeling, University of Oslo. Signe Oksefjell Ebeling has designed a net-based course to go with the text. For comments on the text, I am grateful to my colleagues Kristin Bech and Hilde Hasselgård, University of Oslo. I am also grateful to the students who have attended my courses and whose work I have drawn on in preparing this book.

Oslo, March 2008
Stig Johansson

# Contents

## List of figures

## List of tables

**List of abbreviations**

| | |
|---|---|
| CA | Contrastive analysis |
| CIA | Contrastive Interlanguage Analysis |
| EA | Error analysis |
| ENPC | English-Swedish Parallel Corpus |
| ESPC | English-Swedish Parallel Corpus |
| FL | Foreign language |
| ICLE | International Corpus of Learner English |
| ILS | Interlanguage Studies |
| NICLE | Norwegian component of ICLE |
| OL | Original language (of translation = SL) |
| OMC | Oslo Multilingual Corpus |
| L1 | First language, mother tongue |
| L2 | Second language, target language |
| NP | Noun phrase |
| SL | Source language (of translation = OL) |
| TL | Target language (of translation) |

## 1 Introduction

## 1.1 What is contrastive analysis?

In an interview in *Newsweek* (4 April, 1977, p. 59) Ingmar Bergman, the famous Swedish film director, was reported as saying: 'I still have a tremendous amount of lust to make movies.' As Scandinavians, we know that what he wanted to say was 'har lust att göra film' / 'har lyst til å lage film'.

When we use a foreign language, we may make mistakes because of influence from our mother tongue – mistakes in pronunciation, grammar, and other levels of language – often referred to as *interference*. This is why books in grammar and phonetics for foreign students with a particular mother tongue usually focus a lot on differences between the mother tongue – or *L1* – and the target language – or *L2*.

Being aware of these differences is essential in order to learn the correct and idiomatic use of the foreign language. Without such awareness, we tend to see and hear things in familiar ways, according to the categories which we are familiar with from our native language. And that is not surprising. This is the way we tend to see, hear, and interpret things in general.

Contrastive analysis (CA) is the systematic comparison of two or more languages, with the aim of describing their similarities and differences. CA has often been done for *practical/pedagogical* purposes. The aim has been to provide better descriptions and better teaching materials for language learners. There is more to CA than this, however. When we compare, we often see things more clearly. Some years ago Per Egil Hegge wrote in *Aftenposten* (17 March, 1999):

> Men antagelig er det slik at man ved å se på språkforskjeller, får et skarpere blikk for nyanser. Og et skarpere blikk for nyanser gir en bedre språkfølelse og bedre herredømme over språkets virkemidler […].

Per Egil Hegge was referring to the debate on the value of *nynorsk* as *sidemål* in the Norwegian school. To transfer this to our context: when we compare across languages, we can see the characteristics of each language more clearly, and the comparison can contribute to a better description of each individual language. This type of CA has sometimes been called *analytic comparison* or *linguistic characterology* (Mathesius 1975). Another linguist in the same tradition has put it in this way:

> The contrastive method proves to be a useful heuristic tool capable of throwing valuable light on the characteristic features of the languages contrasted; […]. (Firbas 1992: 13)

To take an example: what does the Norwegian adverb *sikkert* mean? I.e. how certain do you have to be to say *sikkert*? This becomes clearer if we look at it from the point of view of another language. As we shall see later (see **2.7**), *sikkert* is often rendered by *probably* in English (as well as *certainly*, *surely*, etc.).

The importance of CA extends beyond individual languages. When we compare across a number of languages, we can also see more clearly what is characteristic of languages more generally. There is a lot of interest in *universals* of language – that is, what is characteristic of language in general. To study this, there is a need for language comparison. To continue with the example above, we find similar tendencies in the use of expressions of certainty in different languages (**2.7**).

## 1.2 Contrastive analysis and language teaching

The background for CA, as applied to language teaching, is the assumption that the native language plays a role in learning a second language. Mother tongue influence is sometimes very obvious, e.g. in the case of foreign accent. We can often recognise foreign speakers by their accent; an American speaking Norwegian normally sounds quite different from a Frenchman or a German. Mother tongue influence is also apparent in the Bergman quotation above. Influence from the mother tongue is not just negative, however; learning a related language is much easier than learning one that is very different. For example, Håkan Ringbom has shown that Swedish-speaking Finns have a huge advantage in learning English compared with Finnish-speaking Finns (Ringbom 1987). These sorts of observations have probably always been made in language learning and in the contact between native and foreign speakers.

When people have written textbooks for learners of foreign languages, there has regularly been an element of comparison between the native language and the foreign language to be learned. In his book on *Early Contrastive Studies in English*, Thomasz P. Krzeszowski (1995) gives examples that go back to the Renaissance. Bilingual dictionaries are of course also contrastive. But when we refer to CA, we think particularly of a systematic comparison of the mother tongue and the foreign language in order to describe similarities and differences, to identify points of difficulty which might lead to interference. The basic ideas are:

- Describe and compare the mother tongue / L1 / source language and the foreign language / L2 / target language.

- Predict points of difficulty.

- Use the results in order to improve teaching materials.

This sort of approach was developed in the United States in the 1940s and 1950s. Two prominent names were Charles Fries and Robert Lado, who explained the rationale for applied CA in this way:

> The most efficient materials are those that are based upon a scientific description of the language to be learned, carefully compared with a parallel description of the native language of the learner. (Fries 1945: 9)

> The plan of the book rests on the assumption that we can predict and describe the patterns which will cause difficulty in learning and those that will not cause difficulty. (Lado 1957, Preface)

We find a large number of contrastive studies in the 50s and 60s, both in the United States and in Europe, but gradually there was some disenchantment with CA, perhaps because contrastive linguists had made exaggerated claims or because teachers had expected too much. Some problems are (see further Ringbom 1994):

- Only part of the learning problems can be predicted. Many problems are shared, irrespective of the mother tongue.

- Predictions may vary depending upon the linguistic model.

- There is a complicated relationship between difference and difficulty.

- The blinding-flash fallacy: a comparison of L1 and L2 implies that the whole of the two languages get in contact. But the meeting of the languages in the learner's mind depends upon the stage of learning.

The underlying problem is that language learning cannot be understood by a purely linguistic study. So those who were concerned with language learning turned instead to the new disciplines of error analysis, performance analysis or interlanguage studies (see **1.3**), and contrastive analysis was rejected by many as an applied discipline.

## 1.3 Developments

The first stage was the development of ***error analysis*** (EA). The idea was that, rather than predicting points of difficulty, one would observe what problems actually occurred, through a systematic study of learners' errors, thereby revealing learners' difficulties. This information could then be used to improve language teaching. CA could be used as *one* way of explaining errors; this approach was sometimes referred to as the weak claim of CA, as opposed to the strong claim of being able to predict learners' difficulties.

But there are problems also with EA, to do with difficulties of identifying, quantifying, and explaining errors (see Chapter 7). Above all, the observation of errors is not sufficient if we want to identify learners' difficulties, because the most difficult points may be avoided and problems in these areas will not be revealed by EA. So it became clear that the whole of the learners' performance needed to be studied, both errors and non-errors, through ***performance analysis***.

Another development was ***interlanguage studies*** (ILS), i.e. the study of learner language as a system. The aim was to show the gradual development towards the target language. Here the focus is on the learning process, including a comparison with first language acquisition. At this stage we have got quite far from the original idea of CA: perform a linguistic analysis and use this to improve language teaching. The difference in approach could be visualised as in Figure 1.1. CA compares the two languages, while the other approaches are concerned with what language learners actually do in the process of learning.

```
                           CA
                        /      \
                      /          \
                    /              \
L1 ---------------------------------------------------------- L2
              EA / performance analysis / ILS
```
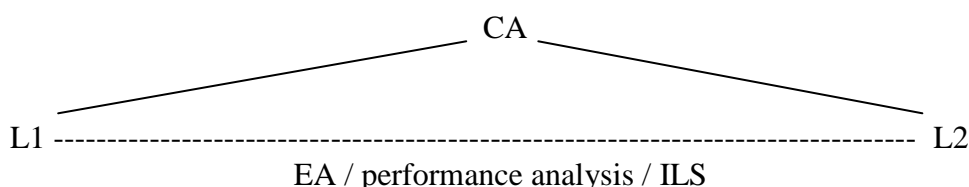
Figure 1.1 A comparison of different approaches

For a while it looked as if applied CA would die out. Some researchers even denied that the native language plays a role in the learning of a foreign language. Such claims are not made any more. The question is not whether mother-tongue influence exists, but *what kinds* there are, *how much*, and *why*. See the exercise at the end of this chapter (**1.13**).

## 1.4 Theoretical and applied CA

As pointed out at the outset, the importance of language comparison extends beyond practical/pedagogical applications and is of great interest in a theoretical as well as an applied perspective. It reveals what is general and what is language specific and is therefore important both for the understanding of languages in general and for the study of the individual languages compared.

In spite of the criticism of applied contrastive analysis, contrastive studies were continued, and their scope was broadened. Although Lado (1957) included a comparison of cultures, early contrastive studies focused on what has been described as ***microlinguistic*** contrastive analysis (James 1980: 61ff.): phonology, grammar, lexis. Examples of research questions:

● What are the consonant phonemes in languages X and Y? How do they differ in inventory, realization, and distribution?

● What is the tense system of languages X and Y?

● What are the verbs of saying in languages X and Y?

With the general broadening of linguistic studies in the 1970s and 1980s, contrastive studies became increasingly concerned with ***macrolinguistic*** contrastive analysis (James 1980: 98ff.): text linguistics, discourse analysis. Examples of research questions:

● How is cohesion expressed in languages X and Y?

● How are the speech acts of apologizing and requesting expressed in languages X and Y?

● How are conversations opened and closed in languages X and Y?

When questions of this kind are raised, it becomes increasingly important to base the contrastive study on authentic texts. This is where multilingual corpora come in. But first we need to touch on the notion of equivalence.

## 1.5 The problem of equivalence

One of the most vexing questions in contrastive analysis is the problem of equivalence. How do we know what to compare? It is not sufficient to contrast formal categories. What is expressed in one language by, for example, modal auxiliaries could be expressed in other languages in quite different ways. Then we do not get very far by a comparison of modal auxiliaries.

One approach is that outlined by Andrew Chesterman (1998: 54) in his proposal for a methodology for contrastive functional analysis:

1. Primary data: instances of language behaviour in different languages.
2. Comparability criterion: a perceived similarity, of any kind, between a phenomenon X in language A and a phenomenon Y in language B. For a given contrastive analysis, this criterion is then defined operationally in terms of a constraint of relevant similarity.

3. Problem: what is the nature of this similarity?
4. Initial hypothesis: that X and Y are identical.
5. Test: on what grounds can the initial hypothesis be supported or rejected? On what conditions (if ever) does it hold?
6. Revised hypothesis (if the identity hypothesis fails): that the relation between X and Y is such-and-such; or, that the use of X and Y depends on such-and-such conditions.
7. Testing of the revised hypothesis.

And so on.

According to Chesterman, the initial hypothesis of identity has the same status as the null hypothesis in experimental studies. The researcher sets out to reject it, but the main point is to show how the perception of similarity is gradually refined in the process of testing.

Translation is a source of perceived similarities across languages. Most linguists working in the field have either explicitly or implicitly made use of translation as a means of establishing cross-linguistic relationships. According to Roman Jakobson (1959: 234), "[n]o linguistic specimen may be interpreted by the science of language without a translation of its signs into other signs of the same system or into signs of another system", and he continues: "Any comparison of two languages implies an examination of their mutual translatability […]". In his book on contrastive analysis Carl James reaches the conclusion that translation is the best basis of comparison:

> We conclude that translation equivalence, of this rather rigorously defined sort [including interpersonal and textual as well as ideational meaning],[1] is the best available TC [*tertium comparationis*] for CA [contrastive analysis]. (James 1980: 178)

As translation shows what elements may be associated across languages, it is fruitful to base a contrastive study on a comparison of original texts and their translations. At the same time, translations have special characteristics of their own and may deviate from original texts in the target language. Given an appropriate corpus model it is, however, possible to control for translation-specific features (see Chapter 2).

## 1.6 Translation studies

Contrastive analysis and translation studies are separate though related fields. Translation studies include a range of approaches, from theoretical and descriptive research on translation phenomena to applied concerns, such as translation assessment and translator training. In one way, it is therefore wider than contrastive analysis. In another, it is narrower, as it is limited to texts that are translated and to the ways in which they are transmitted across languages. In spite of the different aims, there are mutual relationships:

> The relationship between CA and translation is bidirectional. On the one hand, the translation of specific pieces of text may provide the data for CA […]. On the other, CA may provide explanations of difficulties encountered in translation […]. (Hoey and Houghton 1998: 49)

According to Toury (1980: 29), "an exhaustive contrastive description of the languages involved is a precondition for any systematic study of translations". But how can CA at the

---

[1] These types of meaning correspond to the three metafunctions defined by Halliday (see Halliday 2004: 29f.).

same time use translational data as a starting-point and provide explanations of translational phenomena? A solution is again to be found in the choice of appropriate corpus models.

## 1.7 A corpus-based approach

Parallel texts have been used more or less systematically for linguistic purposes since long before the age of computers. A famous early example is the Rosetta Stone, discovered by a young French officer in 1799 in Rosetta, a small town near Alexandria. The stone, now on display in the British Museum, contains inscriptions in three distinct scripts: Egyptian hieroglyphs, demotic script (a late cursive form of hieroglyphs), and Greek. A comparison of these texts eventually led to the deciphering of the hieroglyphs.

If we move closer to our own time, we find that parallel texts in different languages have been used both in translation studies and in comparative language studies. In recent years there has been a fast increasing interest in the use of multilingual corpora in electronic form, i.e. text collections in two or more languages that have been put together in a principled way for language studies. The use of such corpora, with a variety of texts and a range of translators represented, increases the validity and reliability of the comparison. It can be regarded as the *systematic exploitation of the cross-linguistic intuition of translators*, as it is reflected in the pairing of source and target language expressions in the corpus texts. What is new is not the use of parallel texts for language studies, but the fact that multilingual corpora are now being compiled in a systematic manner and are prepared for search and analysis by computer.

## 1.8 Types of corpora

Different types of corpora have been developed for use in contrastive studies (see Chapter 2). Apart from multilingual corpora, we have corpora of learner language (see Chapter 7). The two main corpora which we are going to refer to in this text are the English-Norwegian Parallel Corpus (ENPC) and the Norwegian component of the International Corpus of Learner English (NICLE).[2] At the end of this book (**10.4**) we shall see how the two types can be combined in the new integrated contrastive model.

In addition to these corpora, references will occasionally be made to the LOCNESS native speaker corpus (see **7.9**) and to the Oslo Interactive English Corpus (OIE). OIE is a monolingual English corpus, with exercises developed for a range of problematic areas for Norwegian learners of English.[3]

## 1.9 Correspondence vs. equivalence

Even though we speak of translation equivalence (**1.5**), it is clearly the case that there are different types or degrees of equivalence (Kenny 1998). What we observe in the corpus are correspondences (**2.5**, **2.6**), and we use these as evidence of cross-linguistic similarity or difference or as evidence of features conditioned by the translation process. Analysing the correspondences we may eventually arrive at a clearer notion of what counts as equivalent across languages.

---

[2] For a description of the ENPC, see **2.2** and http://www.hf.uio.no/ilos/forskning/forskningsprosjekter/enpc/. The International Corpus of Learner English (ICLE) and its Norwegian subcorpus NICLE are described in **7.7**-**7.8**.
[3] See http://www.hf.uio.no/ilos/OIE/html/TheCorpus.html.

## 1.10 Structure vs. use

Early contrastive studies typically contrasted the structures of the languages compared. With the availability of multilingual corpora, we can examine not just the available structures, but how they are used in authentic texts. A study of correspondences will reveal that languages often differ in the ways in which they put similar means to use.

## 1.11 Uses of multilingual corpora

Some of the possibilities of multilingual corpora were formulated in this way by Karin Aijmer and Bengt Altenberg (1996: 12):

● they give new insights into the languages compared – insights that are likely to be unnoticed in studies of monolingual corpora;

● they can be used for a range of comparative purposes and increase our understanding of language-specific, typological and cultural differences, as well as of universal features;

● they illuminate differences between source texts and translations, and between native and non-native texts;

● they can be used for a number of practical applications, e.g. in lexicography, language teaching, and translation.

An application which is not mentioned here is natural language processing (see e.g. Véronis 2000, Carl and Way 2003): automatic lexicon extraction, machine-aided translation, information retrieval, etc.

## 1.12 A note on examples and references

Most of the examples in the contrastive chapters are drawn from the English-Norwegian Parallel Corpus (ENPC). Corpus examples are accompanied by a reference code identifying the text, as in:

(1)     His great problem was that he still had no idea what he wanted to do with his life. (RF1)
        Hans største problem var at han ennå ikke hadde noen idé om hva han ville bruke livet sitt til.

(2)     The house was set back from the noisy main road in what seemed to be a rubbish tip. (DL2)
        Huset lå litt tilbaketrukket fra hovedveien, midt i noe som minnet om en søppelfylling.

Here RF1 stands for Robert Ferguson, text 1, *Henry Miller – A Life*, and DL2 for Doris Lessing, text 2, *The Good Terrorist*. The reference code accompanies the excerpt from the original text, and this is generally quoted first.[4]

---

[4] For a full list of corpus texts, see the ENPC website (given in note 2).

**1.13 Exercise**

**The passages below are taken from the International Corpus of Learner English (ICLE). Which of these texts do you think was written by a Norwegian student? How can you tell?**

**Topic**: Some people say that in our modern world, dominated by science, technology and industrialisation, there is no longer a place for dreaming and imagination. What is your opinion?

# Essay 1

Today a lot of people say that it is no longer place for imagination and dreaming. Someone say that science technology steer our lives. I do not quite agree in this assertion, because industrialisation does also make the everyday much easier for us and gives us more sparetime as well.

Advanced technology has made it possible for us to choose where and how we want to live our lives. Some maybe wants to live up in the mountains and some in the country. With modern computers (you can even carry them) and Internet we can take our job with us wherever and whenever we want to. Of course you can not do this in every job, so it depends on what you are doing for your living. In professions like advocate, teacher, doctor etc. among others data is a very helpful invention. You can take your work with you at home, and then choose when you want to do it. A lot of people dreams about living on the countryside or a place which maybe is far from their work. Today these dreams can for many of us easily be fulfilled.

# Essay 2

Science technology and industrialisation are dominating our modern world, but dreaming and imagination have as well place in our life. The scientific developments which have been carried out during the last decades have made life much easier and comfortable to live. The efect of industrialisation have produced a lot of advantages with respect to the quality and safety of the human's life. One of the main advantages that technology offers us is that almost everything is made by machines facilitating the man's work. But, what about the disadvantages of modern world?

In this world almost ruled by computers, full of hate, bigotry, and selfishness, where the best is supposed to be the one who owns most, there is a space for dreaming and imagination. We can not limit our life making use of what modern civilization is offering us, material things which have already been created and that do not let us contribute in personal way. We should stimulate ourselves by doing something new, something which adduce us a development as human beings, creating and making true our dearest dreams. Human nature feels the necessity of escaping from the real life. If we are capable of fergeting about the reality which is surrounding us, we could fly to another worlds which could offer us satisfactions without material things. But sometimes is difficult to cast aside the mask of civilization because it dominates our daily life.

## Chapters 2-6 draw heavily on material published elsewhere and have been omitted from this manuscript for reasons of copyright.

See particularly the following publications:

Johansson, Stig. 1998. On the role of corpora in cross-linguistic research. In S. Johansson and S. Oksefjell (eds), *Corpora and Crosslinguistic Research: Theory, Method, and Case Studies*. Amsterdam: Rodopi.1-24.

Johansson, Stig. 2007. *Seeing through Multilingual Corpora. On the use of corpora in contrastive studies.* Amsterdam: John Benjamins Publishing Company.

## 7. Introducing the study of learner language

### 7.1 Introduction

The study of learner language grew out of error analysis (EA), which had its heyday in the 1960s and 1970s (**1.3**). Rather than predicting learning problems by contrastive analysis (CA), the idea of EA is to observe what problems actually occur, through a systematic study of learners' errors, and use this information to improve language teaching. A systematic study of errors is important not just to identify immediate learning problems, but also because it provides a window into the mind of the language learner.

      The study of errors alone turns out to be insufficient (see **7.5** below) and must be widened to encompass the total performance of learners. We start, however, with some comments on errors in learner language. In order to study errors systematically it is necessary to collect samples of learner language. The next task is to identify the errors. This is by no means easy. We know that language varies a great deal; there is regional, social, and stylistic variation. Native speakers may disagree about what should be regarded as an error. What standard should we use in identifying errors? If we do not have some sort of standard, we would have to accept *I says* and *he done it*, to take just a couple of examples, because these are forms which are actually used by native speakers of English. So we must agree on a norm.

      Perhaps the best way of tackling this problem is to adopt the method described later in this chapter (**7.9**). Here the performance of learner groups is held up against what native speakers of English have done in a similar situation.

### 7.2 Types of error

In the introduction to EA by Carl James, there is a four-way classification of types of error (James 1998: 81):

- *slips*, which can be corrected by the learner himself/herself;

- *mistakes*, which can be corrected by the learner if the deviance is pointed out;

- *errors*, which cannot be corrected by the learner himself/herself without further learning;

- *solecisms*, which involve the infringement of prescriptive rules of correctness, e.g. split infinitives and dangling participles.

The first and fourth types also occur among native speakers. James makes the point that the last type may be characteristic of native speakers, as learners are typically exposed to language where breaches of prescriptive rules are rare. Mistakes and errors are the most important types in connection with foreign-language teaching. But in practice it may be difficult to distinguish between the two categories. In the rest of this chapter we will therefore simply talk about errors.

### 7.3 Causes of error

Errors may occur for a number of reasons. Many errors are **interlingual** (lit. 'between languages', in this case between the mother tongue and the target language) and result from influence from the mother tongue. *Informations* and *advices* (for *information* and *advice*) and *the society* and *the nature* (for *society* and *nature* used in a general sense) are examples we

often find among Norwegian learners of English. Another example of an interlingual error is found in *I still have a tremendous amount of lust to make movies* (cf. **1.1**). Mother-tongue influence is clearly important, but it is not the only source of error.

Another kind of error can be described as ***intralingual*** (lit. 'within language', in this case within the target language). Such errors are a result of overgeneralisation from what has already been learnt. Examples are regularised forms of nouns and verbs, e.g. *sheeps* and *payed* (for *sheep* and *paid*). Paradoxically, errors may also be a result of language teaching, or ***transfer of training***. An example is the overuse of the progressive aspect, which may have something to do with inadequate teaching; see Johansson and Stavestrand (1987).

Often it is impossible to identify one source of error. For example, the use of the progressive may be problematic due to mother-tongue influence, since this is a new category compared with Norwegian. On the other hand, the overuse of the progressive is intralingual in that it reflects what has already been learned and has been overgeneralised. And the problem could be partially teaching-induced.

## 7.4 Communication strategies

Many cases of deviance may be put down to communication strategies, which are attempts at communication when the learner lacks the necessary resources. Examples can be found in the following text quoted from Færch *et al*. (1984: 155). This is a dialogue between a native speaker of English (NS) and a Danish learner (L):

| | |
|---|---|
| NS | do you erm how do you go to school – – erm do you go to school by bus |
| L | |

| | |
|---|---|
| NS | – |
| L | yes sometimes er sometimes I take my er – er – what's it called – er |

| | |
|---|---|
| NS | er your                  oh       what does it look like – |
| L | – er – my – cykel – er (laugh) "knallert" – [ˈknælə] – er (laugh) |

| | |
|---|---|
| NS | (laugh) |
| L | you know er Puch – kn Puch – (laugh) you know so – er some people |

| | |
|---|---|
| NS | mm – (laugh) |
| L | er have a cykel      (laugh) er – – no I can't explain it – you know |

| | |
|---|---|
| NS | aha           aha yer |
| L | some people have a car –     and some people have a bicycle – |

| | |
|---|---|
| NS | oh a |
| L | and some people have a er – erm – a cykel there is a m motor – |

| NS | bicycle – with a motor – | | no no it's – a |
|---|---|---|---|
| L | | motor – is that a bicycle – | oh |

| NS | motorcycle – yer – so you have a bi-cycle I mean or a bicycle and a |
|---|---|
| L | bicycle yer er |

| NS | motorcycle – | (laugh) aha – and you you have a motorcycle – | |
|---|---|---|---|
| L | | oh yer – | yer … |

The problem for the learner is that she does not know the English word for *knallert* ('moped'). Rather than giving up she tries different strategies, sometimes relying on the mother tongue. She borrows from the mother tongue (*cykel, knallert*), she modifies the pronunciation of *knallert* in the direction of English, and she uses the brand name *Puch*. In other words, mother-tongue influence may be due to interference or it may be the result of a communication strategy. According to Færch *et al.*, the most successful strategies are based on the learner's current knowledge of the target language, as they "offer the greatest scope for making creative use of one's communicative resources" (p. 158), though "extensive use of paraphrasing and restructuring strategies may make considerable demands on the addressee's patience" (p. 157).

## 7.5 The insufficiency of error analysis

If we want to understand the difficulties of language learners, it is not sufficient to focus on errors. As pointed out in Chapter 1 (**1.3**), the most difficult points may be avoided and problems in these areas will not be revealed by EA. So we need to study the total performance of language learners and match it against what native speakers do in a similar situation. Through a systematic study, we get an insight into the learners' *interlanguage*, the emerging linguistic system developed in the course of learning. This is in a way 'between' the mother tongue and the target language.

The object of learner language research is to investigate characteristic features of the learner's interlanguage. The question is not just 'Where does the learner go wrong', but also 'What does the learner get right?' and, most important, 'What is the nature of language learning?' and 'How can we best help learners on their way to the target language?'.

## 7.6 Some characteristics of learner language

The learner's interlanguage can be briefly described as a formally and communicatively reduced form of the target language. Some typical characteristics of learner language are:

● it is transitional and changeable;

● there is greater variability than in other language systems;

● there is no speech community and no social norm.

Although there is a lot of variability, features tend to get *fossilised* or fixed, stopping short of full mastery of the target language. Another typical feature is *backsliding*, i.e. slipping back to

earlier stages when the period of active learning stops and the language is insufficiently used. For all of these reasons, the description of learner language represents a great challenge.

## 7.7 The International Corpus of Learner English

Ideally, we would like to study real communication, in both speech and writing, as it develops over time among different groups of learners. In practice, we must be satisfied with less. The most promising development in recent years is the International Corpus of Learner English (ICLE) project, initiated and co-ordinated by Sylviane Granger, Université catholique de Louvain. Essays written by foreign language learners have been collected systematically from a number of mother tongue groups. The essays are chiefly argumentative, on topics generally chosen from a list provided. The number of essays is 250-350 for each mother tongue group, or about 200,000-250,000 words in all. At present (2007) there are sixteen complete ICLE corpora, available in electronic form on CD.[5]

In order to ensure comparability across the different corpora, they had to conform to a set of principles, as shown in Figure 7.1. The learners are all university undergraduates, usually in their twenties. They have studied English as a foreign language. The proficiency level is intended to be 'advanced', although in practice there are considerable differences in proficiency level. The medium is writing, the genre academic essay, the text length between 500 and 1,000 words.

International Corpus of Learner English

Shared features — Variable features

Learner variables | Task variables | Learner variables | Task variables
- Age
- Learning context
- Proficiency level

- Medium
- Field
- Genre
- Length

- Gender
- Mother tongue
- Region
- Other FL
- L2 exposure

- Topic
- Task setting
  - timing
  - exam
  - reference tools
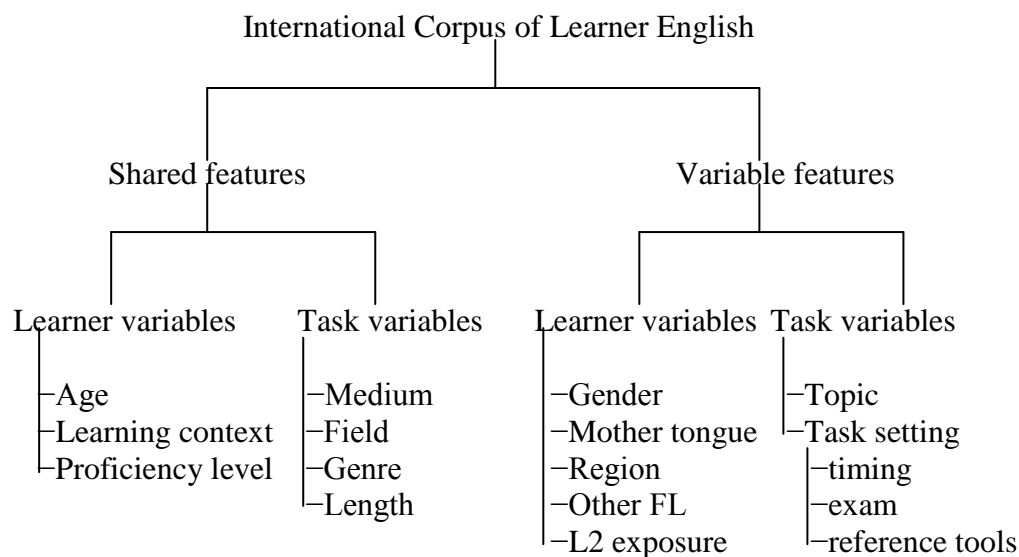
Figure 7.1 ICLE learners and task variables (FL = foreign language, L2 = target language, i.e. English in this case)

Information was collected on a number of variable features, as shown in the figure. The great majority of the learners in all the ICLE subcorpora are female, perhaps because there were more female students available or because these were more willing to take part in the project). Among the mother tongue backgrounds are: Bulgarian, Czech, Dutch, Finnish, French, German, Italian, Norwegian, Polish, Russian, Spanish, Swedish. Other variable features are the learners' country of origin, knowledge of other foreign languages, and the amount of time spent in an English-speaking country. Variable features with respect to task are topic and task setting. There is a wide range of topics, including:

---

[5] See http://cecl.fltr.ucl.ac.be/Cecl-Projects/Icle/icle.htm

- Abortion: right or wrong?
- All armies should consist of entirely professional soldiers: there is no value in a system of military service.
- A man/woman's financial reward should be commensurate with their contribution to the society they live in.
- Crime does not pay.
- Europe.
- For and against breakfast.
- The role of censorship in western society.
- Feminists have done more harm to the cause of women than good.
- Marriage – an anachronism
- In his novel Animal Farm, George Orwell wrote "All men are equal, but some are more equal than others". How true is this today?
- In the 19th century Victor Hugo said: "How sad it is to think that Nature is calling out but humanity refuses to pay heed." Do you think this is still true today?
- Marx once said that religion was the opium of the masses. If he was alive at the end of the 20th century, he would replace religion with television.
- The prison system is outdated. No civilised society should punish its criminals: it should rehabilitate them.
- Most university degrees are theoretical and do not prepare students for the real world. They are, therefore, of very little value.
- Some people say that in out modern world, dominated by science, technology and industrialisation, there is no longer a place for Dreaming and Imagination.

Although the great majority of the topics are on argumentative themes, other types of essays are also represented, including literary essays, e.g.:

- What would you say is the main theme of James Joyce's "The Dead"?
- Arthur Miller's "Requiem" – Hopeful or tragic?
- Fredrick Jackson Turner's Frontier Thesis

Most of the essays are untimed and written at home, with no time limit and with the possibility of using reference tools, such as dictionaries. Timed essays are usually written in an exam situation, where reference tools are not allowed.

In using the ICLE CD we can search for essays that conform to particular criteria, e.g. essays on a particular topic, essays written by male vs. female students and by students with different mother tongues, or a combination of different criteria. For example: essays on "Crime does not pay" by male students with Norwegian as a mother tongue.

## 7.8 The Norwegian subcorpus

The Norwegian subcorpus of ICLE, which we christened NICLE, was compiled in 1999-2002 by Lynell Chvala and Stig Johansson, University of Oslo. The total size of the material is 300+ essays, approximately 212,000 words in all. Most of the learners were first-year students enrolled in English courses at the University of Oslo, but students at other universities/colleges in Norway are also represented. Most of the essays contain 500-700 words. The gender distribution of the learners is 236 female and 82 male students, with the same bias as in the other ICLE corpora.

NICLE is available on the ICLE CD, which can only be accessed from computers connected to the network at the University of Oslo (or other institutions which have a licence

to use the material). In addition to the texts, there is an interface for searching the corpus. Note, however, that such searches are not unproblematic, as the material contains numerous errors (see Chapter 8).

## 7.9 A native reference corpus

The research team on learner corpora at Université catholique de Louvain has also collected a reference corpus of native English essays. This corpus, which is called LOCNESS, is made up of:

- British pupils' A-level essays: 60,209 words
- British university students' essays: 95,695 words
- American university students' essays: 168,400 words

The total number of words is 324,304.[6]

The reference corpus is important in that it provides a standard against which the learners' performance can be matched. In the next couple of chapters there will be examples of studies using different ICLE corpora, sometimes in combination with the LOCNESS reference corpus.

---

[6] See further http://www.fltr.ucl.ac.be/fltr/germ/etan/cecl/Cecl-Projects/Icle/locness1.htm

## *8. Lexical characteristics of learner language*

### 8.1 Introduction

Lexical characteristics of learner language include clear errors as well as other features, such as lack of variation and overuse of common words, which are not wrong but frequently mark the texts as produced by language learners. We start with a survey of some common error types.

All the examples in the chapter are taken from the NICLE material. Many sentences have been abbreviated. In giving illustrations it is a problem that a sentence may contain multiple errors. In some cases errors which are not in focus have been corrected. In other cases, a corrected version is given within square brackets for each portion which is given in italics. The relevant errors are given in bold italics. Usually only the errors in focus are commented on.

### 8.2 Lexical errors

Lexical errors involve the misuse of individual words in the target language. They may be wrongly spelled, e.g. *descent* for *decent* or *wether* for *whether*. There may be other formal mistakes in the use of words, as in:

(1)     Meeting different kinds of people is a very important part of our ***lifes*** [lives].

(2)     As teaching *is, as much of* [is just as much] ***a practical work*** [practical work] as ***an intellectual work*** [intellectual work] you learn to be a teacher by being a teacher.

(3)     If they ***fail getting*** [fail to get] a job, they often do what they know they do best, criminal stuff.

(4)     In America, they have put up posters outside pedophiles' houses to ***tell that*** [tell/inform people that] here lives *an* [a] criminal.

In (1) the noun *life* is given the wrong plural form. In (2) *work* is used as a countable noun in a way which is not possible in English. In (3) the verb *fail* is followed by an *ing*-form rather than an infinitive. The use of *stuff* in this example is not very precise, but it cannot be categorised as a clear error. In (4) *tell* is used without an indirect object, which is normally required with this verb.

In the examples given so far, it has been possible to give a correction without choosing another content word. The focus below will be on errors involving the wrong choice of content words.

### 8.2.1 Equivalence errors

Many errors are interlingual and arise because a word in the target language is wrongly equated with a word in the mother tongue. Examples of such equivalence errors are:

(5)     There are a lot of examples where the teacher is very intelligent, and knows a lot, but he is not able to ***learn*** [teach] the children what he knows.

(6)     This program would also make the inmates feel useful, and ***learn*** [teach] them that crime is not the solution to their problems.

Here *learn* is equated with *lære* in Norwegian, which corresponds both to *learn* and *teach*. Another verb which is misused by Norwegian learners is *mean* in examples like:

(7)     I know that many of the teachers at the College **mean that** [think] they do not have the time to teach us these things.

(8)     Some people **mean that** [think that] the feminists have done more harm to the cause of women than good.

The learner here uses English *mean* in the same way as Norwegian *mene* to express the meaning 'think, be of the opinion'. But the two verbs may correspond in other cases, e.g. in *What do you mean by that?/Hva mener du med det?* Other examples of equivalence errors are:

(9)     In professions like **advocate** [cf. advokat; lawyer], teacher, doctor etc. among others **data** [the computer] is a very helpful invention.

(10)    Have you ever thought of how many people *who're* [are] working in the **data** [cf. data; computer] business?

(11)    The police **uphold** [cf. opprettholde; maintain] order by their presence in the streets, but it's the **arrestations** [cf. arrestasjoner; arrests] and jailing of criminals that really rids *the society* [society] of its unwanted.

(12)    It is said that all the movies and computer games ruin the child's **fantasy** [cf. fantasi; imagination].

(13)    People who believe that **fosters** [cf. foster; embryos] are human beings will probably say it is wrong to have an abortion no matter what.

(14)    Some don't even prepare their **lections** [cf. leksjon; lessons] so you can't *follow on* [follow] what they are saying.

(15)    If the parents are very young, they might not have an **economy** [cf. økonomi; a financial situation] that is **stabile** [cf. stabil; stable] enough to have a baby.

Note also that example (9) is unnecessarily verbose; *like* alone would do in giving examples of professions, *etc.* and *among others* could be left out. In (10) *who* reflects the use of *som* in such constructions in Norwegian (*hvor mange mennesker som …*); we could perhaps call this a grammatical equivalence error (for such grammatical errors, see **9.1**). These are some more examples of lexical equivalence errors:

| Error | Norwegian model | Correction |
|---|---|---|
| argument for | argumentere | argue |
| fantasise | fantasere | imagine, use one's imagination |
| realise | realisere | carry out, implement |
| [the time they] spare | spare | save |
| spring up from | springe ut av | have its origin in |
| [receive] critics | kritikk | criticism |
| datagames | dataspill | computer games |
| examinator | eksaminator | examiner |

| | | |
|---|---|---|
| inspirator | inspirator | inspirer |
| lonely-mother | alenemor | single/lone mother |
| Middle Europe | Mellomeuropa | Central Europe |
| theoretics | teoretikere | theoreticians |

Some equivalence errors involve the transfer of idioms and longer expressions, as in:

(16) ***My meaning with this is*** [cf. min mening er; I think] that even if some feminists have taken it *to* [too] far, they can never do more harm to the cause of women than all the good things other feminists have done.

(17) Even if ***the lawbreaking was*** [cf. forbrytelsen var; the crime was committed] long ago and he regrets it, it often ***has nothing to say*** [cf. har ikke noe å si; does not mean anything, is of no consequence].

(18) One of the good things *of* [about] being a human being is that we *have the possibility to* [better: can] ***put word on*** [cf. 'sette ord på'; express in words] what we desire.

(19) Nothing is so important as time. If you ***have time on your side*** [cf. 'har tiden på din side'; time is on your side] you will always prevail.

(20) Many people have *made* [developed] great theories ***on the paper*** [cf. 'på papiret'; on paper].

(21) Sometimes theoretical people cannot imagine how things will be once they are ***put in to life*** ['sette ut i livet'; put into practice].

(22) It may ***say something about*** [cf. 'si noe om'; be significant for] the way we perceive *ourself* [ourselves].

Lexical equivalence errors are very common, but we also find many intralingual errors which reflect incomplete learning of the target language.

## 8.2.2 Conceptual confusion

Where there is conceptual confusion, students have learned a word in the target language, but confuse it with a semantically related word. Examples:

(23) The most important ***issue*** [problem] to help prisoners with is to help them understand that they cannot keep on ***disobeying*** [breaking] the law.

(24) One of the differences from "the good old days" *who* [that/which] *attracts* [attract] people is that the television is showing the ***watchers*** [viewers] what the generations before the 20th century had to imagine.

(25) A better solution would have been if the colleges and the universities ***had practical exercise*** [offered practical exercises/training]. That would give the students who are weaker in theoretical terms a chance to show what they *can* [can do].

(26) If you want to earn a lot of money you ***educate*** [train] *for becoming* [to become] for example a lawyer instead of a teacher or a nurse.

Conceptual confusion overlaps with problems of collocation (**8.6**), e.g. in the case of *disobeying the law* in (23). Often we find both conceptual confusion and other types of error in the same sentence. In (24) the choice of relative pronoun is wrong. In (26) *for becoming* reflects the grammar of *for å bli*.

As a final example, consider this passage from an essay discussing whether university degrees are theoretical and do not prepare students for the real world:

(27)    Many critics of *the university degree* [university degrees] focus their criticism on *wether* [whether] the knowledge ***achieved*** [acquired] at the universities *are* [is] too general and in many cases not *usefull* [useful] for the *work* [jobs] that most companies can offer. They point to "høgskolen" and "yrkesskolen" to ***understate*** [underpin] their argument. They, "høgskolen" and "yrkesskolen", represent two institutions which focus their education on ***concrete*** [practical] jobs.

Here there is a mixture of problems ranging from errors in spelling and grammar to conceptual confusion. As used here, *understate* is almost the opposite of the definition we find in the *Longman Dictionary of Contemporary English*: 'to describe something in a way that makes it seem less important or serious than it really is'. Possibly the error can be put down to a formal confusion with *underline*.

## 8.2.3 Formal confusion

In some cases it is obvious that the confusion of words is purely on the level of form, as in:

(28)    Today you will observe more and more ***confirmity*** [conformity] *with* [among] women. Personal identity and individuality are disappearing.

(29)    The modern super model ***carrier*** [career] woman doesn't need a man except as a sex partner.

(30)    But who can we ***blain*** [blame], and is this the reality of our society? Has the world really come to an end? To quickly answer the questions stated, the only persons we can ***blain*** [blame] *is* [are] ourselves.

(31)    Idealistic struggles when ***loosing*** [losing] objectives to fight for too often *turns* [turn] into ***exstremeties*** [extremes], and so also with the western *feminists* [feminists'] fights.

(32)    It is a subject that has been discussed over and over again; still there is no ***resolution*** [solution] in sight.

(33)    Being surrounded by other criminals with negative *attitude* [attitudes] will ***effect*** [affect] your mentality.

(34)    This has later been proven wrong, but at the time it was introduced people fell for it, ***exepting*** [accepting] the war and making the process of the war easier for the authorities.

There is no semantic connection between *confirm* and *conformity* in (28) or between *carrier* and *career* in (29). In (30) the confusion seems to be purely phonetic, since there is no such word as *blain*. Though there is an etymological connection in the next three examples, the

meanings of *resolution* and *solution*, *effect* and *affect*, and *accept* and *except* are quite distinct. The formal confusion in this example is more complex:

(35)    If we want to *scare* [discourage] criminals from committing their deeds, we can *explain our **brethren and sisthren*** [to our brothers and sisters] what would happen if everybody did what they do.

Again, there are several errors, but the one that is most relevant in this context is *brethren and sisthren*. *Brethren* is a possible plural form, but in present-day English it is generally used about members of a religious group. Here it has also given rise to the non-existent form *sisthren*.

## 8.2.4 Formal-conceptual confusion

Sometimes there are elements of formal as well as conceptual confusion, as in (36), where the verbs *lie* ('ligge') and *lay* ('legge') are mixed up:

(36)    They don't **lay** [lie] on the lawn and watch the clouds and dream themselves away.

Such mix-up is also found in the case of *rise* ('stige') vs. *raise* ('heve'). Formal-semantic confusion is probably also behind the use of *aspect* for *respect* in examples like:

(37)    Therefore it is possible to state that in this **aspect** [respect] opium and television *is* [are] comparable.

(38)    Of course, television cannot replace religion in all **aspects** [respects].

Formal-semantic confusion may lead to forms which look very strange indeed:

(39)    Personal interests like greed and **gormandision** [gluttony] will make us cross the border to crime without regard *to* [for] our personal environment.

There is no such word as *gormandision* in English. The closest are *gourmand*, which means someone who likes to eat and drink a lot, and *go(u)rmandize* 'to eat like a glutton'. Here we are on the borderline of the next category.

## 8.2.5 Word formation errors

Errors may also arise through the incorrect use of English word-formation patterns, as in:

(40)    The **suspected** [suspect] is entitled to defend him/herself, s/he will not be punished before the *trail* [trial] is over, and the person is found guilty.

(41)    Many are **judgementive** [judgemental] in their way of thinking that rich people are better than those who haven't got money.

(42)    Besides *this* [these] achievements, the woman has gained **understandness** [understanding] and respect from men, as this is what the women's fight *concerns about* [is about, concerns, is concerned with].

In the last example we also note the error in the grammatical context of *concern*. We can be *concerned* (i.e. worried) *about something* or a matter can *concern something* or *be concerned with something*.

## 8.2.6 Colligation errors

Colligation is the term used for the bond between words and their grammatical context. Examples (3) and (4) at the beginning of the chapter illustrate errors in the complementation of the verbs *fail* and *tell*. In example (35) above the verb *explain* requires a following *to*. Other examples of colligation errors are found in:

(43)    I do not **remember to have seen** [remember having seen] the six of them comment on anything else than looks and social status.

(44)    Some people spend so much time in front of the television that they have almost **stopped to communicate** [stopped communicating] with other people.

(45)    Some people *mean* [think] *its* [it's] a waste of money to help the criminals with *teraphy* [therapy], while other people *mean* [think] that a criminal should be **treated** [treated as] equal with any other person.

(46)    Why on Earth shouldn't the raising of children be **regarded** [regarded as] equally important as some pathetic career in an advertising agency, selling detergent.

As for the choice of complementation pattern in (43) and (44), see Hasselgård *et al*. (2004: 370f.). In (45) and (46) the verbs *treat* and *regard* require a following *as*; note, incidentally, that *regard* and *consider* are different in this respect.

Students are strongly recommended to make regular use of bilingual dictionaries (to prevent equivalence errors) and monolingual English dictionaries (to combat intralingual errors). Note that learners' dictionaries give information not just on form, meaning, and style, but also on colligation.

## 8.3 Overuse and underuse

The vocabulary of learner language is characterised not just by errors but by being less varied when measured against the performance of a comparable group of native speakers. Common words are often overused. The tendency to cling to what is common and familiar has been referred to by the expression 'lexical teddy bears' (**8.5**). The overuse of common words at the same time means that many other words are underused. In our case overuse and underuse are measured by comparing word frequencies in the ICLE corpora and the native speaker reference corpus LOCNESS (**7.9**).[7]

## 8.4 Word frequencies

Ringbom (1998) compared the vocabulary of seven ICLE corpora with the LOCNESS native speaker corpus (**7.9**). As the corpora vary somewhat in size, he compared frequencies per 10,000 words (relative frequencies) rather than absolute frequencies. The study revealed that many words were more frequent in all the ICLE corpora than in the native speaker reference corpus.

---

[7] The terms 'overuse' and 'underuse' may also apply to translations (cf. **2.8**). In this case translations are matched against the standard of comparable original texts.

The groups of words below have been selected from Ringbom's lists, with the addition of corresponding figures for NICLE. We start with two common verbs (only the forms listed are included, not the whole paradigm) and a common sequence:

|  | NS | FRE | SPA | FIN | FINSWE | SWE | DUTCH | GERM | NORW |
|---|---|---|---|---|---|---|---|---|---|
| *get* | 6 | 7 | 18 | 18 | 16 | 16 | 14 | 19 | 26 |
| *think* | 6 | 21 | 21 | 22 | 30 | 30 | 16 | 22 | 27 |
| *I think* | 3 | 10 | 5 | 7 | 16 | 16 | 6 | 9 | 12 |

NS = native speaker corpus, FRE = French, SPA = Spanish, FIN = Finnish, FINSWE = Finland Swedish, SWE = Swedish, GERM = German, NORW = Norwegian.

The following example illustrates how the frequencies per 10,000 words were calculated for *get* in NICLE:

Absolute frequency: 549
Total number of words: 212,727
Frequency per 10,000 words:

$$\frac{549}{212,727} \quad x \quad 10,000 \quad = \quad 26 \text{ (after rounding)}$$

We can see that the two verb forms are used more often by all the learner groups, and the same applies to the sequence *I think*. The use of *I think* can be illustrated by the following extract from a NICLE essay:

> *In my opinion, I don't think* there is much room left for dreaming and imagination in our modern world and society. There are so many obligations and commitments, and we are very much controlled by our daily routines. *I don't think* there is a lack of dreams and imagination within the population, but *I think* that we don't give ourselves time to just let our thoughts wander.

There are no blatant errors here, but the repeated use of *I think* in such a short passage is conspicuous. Besides, the first sentence opens with two expressions of speaker attitude; one would be sufficient. From the comparison of frequencies, we conclude that native speakers have more ways of expressing personal opinion or that they do not intrude into the text to the same extent as the learners.

First- and second-person pronouns in general tend to be overused by all the learner groups:

|  | NS | FRE | SPA | FIN | FINSWE | SWE | DUTCH | GERM | NORW |
|---|---|---|---|---|---|---|---|---|---|
| *I* | 25 | 45 | 36 | 52 | 94 | 88 | 41 | 136 | 87 |
| *you* | 8 | 33 | 33 | 34 | 49 | 31 | 46 | 76 | 61 |
| *we* | 34 | 81 | 98 | 65 | 121 | 120 | 34 | 41 | 90 |

We shall come back to this in connection with reader-writer visibility (**9.4**).

Another consistent difference is found with coordinating conjunctions:

|  | NS | FRE | SPA | FIN | FINSWE | SWE | DUTCH | GERM | NORW |
|---|---|---|---|---|---|---|---|---|---|
| *and* | 227 | 235 | 269 | 276 | 262 | 302 | 274 | 273 | 293 |
| *but* | 36 | 66 | 70 | 57 | 55 | 58 | 64 | 67 | 55 |
| *or* | 42 | 52 | 62 | 56 | 63 | 50 | 46 | 59 | 53 |

We can take this to reflect the more common choice of coordination among the learners, while native speakers prefer more complex structures. We shall come back to *and* and *but* in connection with sentence connection (**9.3.2**).

Two more groups where we find fairly consistent overuse are:

|        | NS | FRE | SPA | FIN | FINSWE | SWE | DUTCH | GERM | NORW |
|--------|----|-----|-----|-----|--------|-----|-------|------|------|
| *so*    | 16 | 27  | 33  | 30  | 27     | 26  | 31    | 31   | 30   |
| *too*   | 6  | 10  | 9   | 12  | 10     | 10  | 12    | 11   | 10   |
| *very*  | 14 | 22  | 31  | 25  | 24     | 16  | 28    | 24   | 17   |
|        |    |     |     |     |        |     |       |      |      |
| *always* | 4 | 11  | 10  | 11  | 11     | 8   | 10    | 11   | 9    |
| *people* | 35 | 62 | 63  | 71  | 63     | 65  | 70    | 57   | 64   |
| *things* | 4  | 6  | 16  | 13  | 9      | 18  | 12    | 12   | 15   |

Though *so* has a number of different functions (intensifier, conjunction, etc.), we can interpret the first group as reflecting the tendency among learners to go for more general intensifiers, while there is more variation among native speakers or less use of intensification. The words of the last group, together with the verb *get* (cf. above), give evidence of a preference among learners for common, well-known words, for which Hasselgren (1994) has introduced the term ***lexical teddy bears***.

## 8.5 Lexical teddy bears

The observations on word frequencies fit in very well with Angela Hasselgren's lexical teddy bears, a metaphorical expression which compares a language learner's preference for familiar words to a child's emotional attachment to a cuddly toy. Hasselgren's article reports on two studies, the first of which is an analysis of translations from Norwegian into English. She distinguishes six routes leading to lexical problems. Three are interlingual (cognates, transliteration, and perceived equivalents; cf. **8.2.1** above) and the other three intralingual (synonyms, association, core items). Synonyms and association are comparable with conceptual and formal confusion (cf. **8.2.2**-**8.2.4**). *Cores* are central concepts with a wide range of use, such as the noun *thing*, the adjective *nice*, the adverb *very*, and the verb *get*. The use of lexical teddy bears may arise both via interlingual and intralingual routes, in the former case drawing on the mother tongue, in the latter case making use of what is familiar in the foreign language.

The routes may lead to ***dissonance*** of different kinds: semantic, stylistic, collocational, etc. The notion of dissonance, rather than error, is significant. If we want to characterise learner language, we must include both clear errors and types of deviance which are more subtle. There may be something about the learner text which does not quite sound right, without involving clear errors, frequently reflecting the overuse of common words, or core items.

In her second study Hasselgren compares reactions to test items by group of learners and a native speaker group. This is a similar type of study to the comparison of NICLE and LOCNESS, except that it involves isolated sentences in a testing situation. The results reveal a tendency among learners to go for familiar words, which they apparently perceive as safe choices, though they may lead to dissonance. For example, the learners were more likely to use core items like *very much* and *a lot* as intensifiers, whereas there was more variation in the native speaker group. Core verbs, such as *get* and *have*, were also chosen more commonly by the learners in combinations of verb plus object, e.g. in a task where they were asked to

supply verbs, as in *He has _____ a reputation as a playboy* and *He has _____ treatment for his ulcer*.

Hasselgren's studies are significant in focusing on less blatant aspects of deviance in learner language. She concludes that learners have a tendency to cling to words which they feel safe with, "perhaps most forcibly in the case of overuse of cores at the expense of specific collocations" (p. 256). The remedy she proposes is to focus on language in context:

> By releasing learners from their conception of words as isolated items, often with roots somewhere in the first language, we may be able to equip them with a true sense of word appropriateness, and the confidence to throw away their lexical teddy bears. (p. 257)

This brings us naturally to the topic of collocations.

## 8.6 Collocations

Collocations are bonds between words which have a tendency to occur together (cf. **3.1**). Many problems in learner language are due to inappropriate collocations, often involving the overuse of common words. Some examples are:

(1)     If they *fail getting* [fail to get] a job, they often do what they know they do best, ***criminal stuff*** [commit crimes].

(2)     In Norway the prisons are quite ***nice*** [pleasant, comfortable] *comparing to* [compared to] prisons in other countries.

(3)     I think television has the ability to *passify the viewer* [make the viewer passive] to a ***big*** [great] extent.

In (1) the informal general noun *stuff* is used rather than a more specific wording. *Nice* in (3) is an adjective which is commonly overused. In (3) we find the informal adjective *big* instead of *great*, which goes better with the abstract noun *extent*. These are some other examples where other adjectives would be more appropriate than *big*:

| | | |
|---|---|---|
| in a big degree | > | to a great extent |
| a big amount of calories | > | a large amount of calories |
| this big international treaty | > | this important international treaty |
| a big financial problem | > | a serious/pressing financial problem |
| big accidents | > | serious accidents |
| big commitment | > | firm/strong commitment |
| a big discussion | > | an intense/vivid discussion |

Verbs are often involved in inappropriate collocations. In (4) the verb *steer* probably reflects the Norwegian cognate *styre*:

(4)     *Someone* [some people] say that science and technology ***steer*** [guide] our lives.

The core verbs *do* and *make* are used inappropriately in:

(5)     In some cases people are so helpless that they go out and ***do*** [commit] a crime just to get *in to* [into] jail so that they can survive.

126

(6)     When a criminal gets sentenced to death, he or she must have ***done*** [committed] a serious crime like a murder.

(7)     The authorities often ***made*** [took] advantage of this trust *to* [in] God and used the gospel *manipulative* [manipulatively] to keep inhabitants calm

(8)     Many people have ***made*** [developed] great theories *on the paper* [on paper].

(9)     As an against-breakfaster you can also ***make*** [do] a good deed by *bying* [buying] a meal in the *cantine* [canteen], and thereby *rise* [raise] their income.

The choice between *do* and *make*, as in (9), is complicated by the fact that both may correspond to Norwegian *gjøre*.

## 8.7 Prepositions

A great many errors involve the inappropriate use of prepositions. This is hardly surprising, because prepositions are common and have many uses. Moreover, prepositions which are naturally associated cross-linguistically, such as *on* and *på*, often do not correspond.

The correct use of prepositions cannot be learned in isolation. Prepositions are often ***bound*** to a preceding word (cf. **6.3**). For example, some nouns and adjectives require a following preposition, and many verbs take a particular complementation pattern: *belief in, confidence in, knowledge of*; *angry with, ashamed of*; *argue about, care for, take care of*, etc. Many errors involve bound prepositions, as in:

(1)     The authorities often *made* [took] advantage of this ***trust to*** [trust in] God and used the gospel *manipulative* [manipulatively] to keep inhabitants calm

(2)     A person will always feel ***guilty for*** [guilty about] what he has done and if he gets caught it is even worse for him.

(3)     People do not learn to be ***critical against*** [critical of] their society by watching television.

(4)     I do ***agree in*** [agree with] that statement.

Learning how to use bound prepositions is a question of mastering sequences.

***Free*** prepositions are not bound to a preceding word, though they often enter into sequences with following words: *in addition, in the same way, in other words*, etc. The same preposition can be both bound and free, depending upon context. Compare:

(5)     I'm prepared ***on*** [for] dying early.

(6)     On *week ends* [weekends] people go shopping, go to cabins ***on*** [in] the *mountain* [mountains], or by the ocean, or in the woods.

In (5) *prepared* has to be followed by *for*, a clear case of a bound preposition; another possible option is *prepared to die*, with a *to*-infinitive. The error is clearly inspired by Norwegian *forberedt på*. In (6) *cabin* does not require a particular preposition, but *on* is used incorrectly, probably reflecting Norwegian *på fjellet*.

Free prepositions typically have a more clearly defined meaning than bound prepositions, such as denoting time (e.g. *during*) or place (e.g. *between*), but abstract and metaphorical meanings are common. Compare:

(7)     Some people believe in UFOs. But we don't know if there is some kind of life *at* [on] other planets, and how that life would be like.

(8)     *At* [in] the historical and philosophical subjects the students go to classes *listening* [and listen] to the professors, they read a lot, and they learn a lot.

In (7) the preposition has a literal, local meaning; in (8) it is used metaphorically. In these examples *at* is wrongly chosen instead of *on* and *in*. The errors here are not interlingual; the wording in Norwegian would require *på* and *i*. Distinguishing *at*, *in*, and *on* is a challenge for Norwegian learners. Note incidentally the additional error in (8); the use of the *ing*-form suggests that the students go to classes while they are listening to the professors, a rather unlikely situation.

     In this context it is not possible to do more than give some examples of prepositional errors in the NICLE material. We start with straightforward formal errors and move on to more complicated examples of preposition choice.

### 8.7.1 Formal errors: prepositions

A preposition which is often given an incorrect form is *into*. Compare:

(9)     You can for example break *in to* [into] a store.

(10)    Almost all Norwegian male citizens are called *in to* serve their country.

In (9) we have the preposition *into*, which should be written without a space. In (10) we have the adverbial particle *in* (part of the phrasal verb *call in*) followed by a *to*-infinitive. This is correct.

     English has a number of multi-word prepositions where learners may choose the wrong form, e.g. in:

(11)    But *in despite of* [despite, in spite of] all this, "the old fashioned" objects have survived.

(12)    *Woman* [women] have *achieved* [reached] a position that gives *her* [them] all basic rights *as regards to* [as regards, with respect to] education.

(13)    The new technology has brought about changes in people's lives *with regards to* [with regard/respect to, as regards] diseases and pregnancy.

(14)    People eat breakfast all over the world, but *except from* [apart from, except for] the fact that the meal is eaten in the morning, there are a lot of differences between the countries when it comes to this meal.

(15)    Karl Marx believed that religion made people passive. And *besides from* [besides, apart from] making people passive it also kept them down in many ways.

(16)    *Likewise to* [like] religion, television makes people passive.

Here we see that similar forms are confused. These errors are, however, a small minority compared with the next category.

## 8.7.2 Errors in preposition use

The list below gives examples of inappropriate preposition use in the NICLE material. As you go through the material, you may consider the following points: Does the error involve a bound or a free preposition? Is the error interlingual, i.e. influenced by a similar form in the mother tongue? Note that in some cases it is necessary to rephrase the sentence rather than just insert a different preposition.

### *Against*
(17)    People do not learn to be critical ***against*** [of] their society by watching television.

(18)    Advanced weapons, *as* [such as] long-range ballistic missiles, armed with toxic gasses, nuclear warheads etc. could surely be said to represent one of the largest threats ***against*** [to] world peace.

### *As*
(19)    I could also list *other ethnic discriminations* [other cases of ethnic discrimination] in our modern community, ***as*** [such as] the blacks in South Africa, the Albanian people in Kosovo and the Lapps in Norway.

Cf. also (18) above.

### *At*
(20)    The problems from all over the world are thrown into your own living room, and after a *wail* [while] the *watchers* [viewers] are getting immune to other *peoples* [people's] suffering, seeing it every day ***at the*** [on] television.

(21)    I now ***attend at*** [attend] Bislet College in order to take an English degree.

(22)    They punish you for example with isolation ***at*** [in] your room.

(23)    You didn't have to worry so much about staying ***at*** [in] good health

(24)    We are ***at*** [on] the very edge of mapping the biological mysteries.

### *Beside, besides*
(25)    I do not think I have seen any programme, ***beside*** [besides] a couple of documentaries, that has criticised *the American society* [American society] or any society for that matter.

(26)    That's ***besides*** [beside] my point.

### *Concerning*
(27)    Former prisoners should receive assistance ***concerning*** [in] finding a place to live

(28)    It can be hard to *define* [draw] the line ***concerning*** [between] whether something should be censored or not.

***During***

(29)   The military system has been changed considerably ***during*** *the recent years* [in recent years].

(30)   ***During*** [in] the last couple of years we have gotten a new genre in television shows called reality TV.

(31)   Many ideologies have been suggested and tried out in different societies ***during*** [over] the years.

***For***

(32)   Some of those have *urge* [an urge] to escape from reality in search ***for*** [of] themselves.

(33)   We also got a great offer to buy a laptop ***for*** [at] a real good price.

(34)   We are not the same as ***for*** [a] thousand years ago.

(35)   We learned that we should be happy ***for living*** [to live] in our own country.

(36)   It may seem that the universities are *useing* [using] the wrong methods ***for getting*** [to get] the students on the right track

(37)   People ***claimed for*** [demanded] shorter days at work and more holidays.

(38)   Marx, and most of the dechristianized *middel* [middle] classes at his time, almost replaced art ***for*** religion [replaced religion by/with art; or: substituted art for religion].

***From***

(39)   The question is whether we, with the new technology, no longer have a place for dreaming and imagination. Does it deprive us ***from*** [of] social contact?

(40)   But ***unlike from*** [unlike] religion, people in today's society are rational *beeings* [beings] with the oppurtunity [opportunity] to choose.

***In***

(41)   Kids are so easy to fool, they ***believe in*** [believe] every word, and *every thing* [everything] they hear and see.

(42)   I do not quite agree ***in*** [with] this assertion.

(43)   Many people are a lot better ***in*** [at] doing things practically.

(44)   Surveys show that television as a phenomenon has a great impact ***in*** [on] most *peoples* [people's] lives.

(45)   Everything happened ***in*** [at] an enormous fast speed.

(46)   The course in Norwegian has the same level *as* [as at] the University ***in*** [of] Oslo, and that is great.

(47)   An idea would probably be to let the students go out ***in*** [into] the real world once a week all through the years of study.

*Of*

(48)    These are my views *of* [on] how we should deal with offenders.

(49)    Gabriel is developing through more insight *of* [into] himself.

(50)    I never went to church, except when I was *Baptised* [baptised], *Confirmed* [confirmed], and took part *of* [in] my father's burial.

(51)    During the *1960ᵗʰ* [1960s] the Civil *Right* [Rights] movement inspired women to try to obtain better *condition* [conditions] through campaigns *off* [of] mass agitation.

*On*

(52)    Trevor chooses to do things he likes *on* [in] his spare time. He likes to travel but has only time to do that *on* [during] his vacations.

(53)    Examples *on* [of] places that would be in the danger zone are: …

(54)    No one should be sent to jail *on* [for] the sole reason that they are drug addicts or abusers.

(55)    Workers sold their "muscles" to the capitalists, who profited *on* [from] the products that were produced and sold.

(56)    *There has been several researches done* [Several research projects have been done] *on* [in] this field.

*Over*

(57)    These are two questions l will try to reason *over* [about] in the following paragraphs.

(58)    The debate *over* [on] the professionality of the army has been going for decades.

*To*

(59)    We are changing *to* [for] the better every second.

(60)    There are not enough jobs *to* [for] everyone.

(61)    My conclusion *to* [on] this is that we just have to accept that the world is changing.

(62)    *According to* [in] my opinion, there will always be a place for dreaming and imagination.

*Under*

(63)    Forgery of money and credit cards, embezzlement and *transactions* [criminal transactions] are some examples of crime, which goes *under* [in] this category.

(64)    They *live under other influences than we did* [are influenced by other conditions than we were, live under other conditions than we did].

*With*

(65)    Our western society prides itself *with* [on] being the land of the free.

(66)    Being a romantic does not always mean that one escapes, in fact most romantics were passionately concerned *with* [about] the state of the world.

## 8.8 Spelling

There are many spelling errors in the NICLE material, as should be evident from numerous corrections in the example sentences above. This not surprising, considering the many irregularities of English spelling. Frequently the errors involve letters which can stand for similar sounds, e.g. *s* and *c*:

*desent*, *descent* for *decent*
*neccessary, nessessary* for *necessary*
*sermonies* for *ceremonies*

Unstressed syllables may be misspelt, as in:

*secondery* for *secondary*
*separete* for *separate*

The ending *-ful* in adjectives is often misspelt with a double *ll*: *forgetfull, meaningfull, shamefull, successfull, usefull*. Words may be influenced by related forms, as in *explaination* for *explanation* (influenced by *explain*).

     Occasionally, a word is spelt according to Norwegian conventions, as in *after a wail* for *after a while*. *Feduses* presumably reflects the American pronunciation of *foetus*. With some less common words, there may be a lot of variation, e.g. *techonology, technologhy, tecnology* for *technology*. Rather than enumerating examples more or less at random, we shall look briefly at three recurrent error types.

### 8.8.1 Word pairs

The following common words are often confused:

- the preposition *of* and the adverb/preposition *off* (*of course*, *think of* vs. *take off*, *off the wall*);
- the possessive determiner *its* and the contracted verb form *it's* (*its truth* vs. *it's true*);
- the possessive determiner *their* and the adverb *there* (*their house* vs. *there is a house there*);
- the relative/interrogative determiner *whose* and the contracted verb form *who's/who is* (*whose is it?* vs. *who's there?*)
- the adverb *then* and the conjunction *than* (*see you then* vs. *better than ever*);
- the preposition and infinitive marker *to* and the intensifier *too* (*to school*, *to be* vs. *too good*);
- the verb *affect* and the verb/noun *effect* (*the disease affects the nervous system* vs. *harmful effects*, *they lack confidence in their ability to effect change or improve the quality of their lives*);
- the verbs *proceed* and *precede* (*proceed along the road* vs. *he preceded the current president*);
- the nouns *price* and *prize* (*the price of the house* vs. *she received a prize*);
- the nouns *desert* and *dessert* (*the Sahara Desert* vs. *we had fruit for dessert*);
- the adjective *principal* and the noun *principle* (*the two principal actors* vs. *a guiding principle*);
- the adjective *loose* and the verb *lose* (*the screw is loose* vs. *lose one's job*).

Accordingly, we find errors such as:

(1)     A child also needs *it's* [its] mother more *then* [than] anything after it's born.

(2)     "One Nation" wants nothing to do with Asians, Africans or anyone *whose* [who is] not from the Western world.

(3)     *Therefor* [therefore] I say that abortion is right, but this issue will never be resolved.

(4)     Millions of people would *loose* [lose] their jobs when the cereal factories would have to close down.

(5)     These are very important *principle* [principal] rights that establish the women's rights to be human *beeings* [beings] on the same level as men, and are basic rules in *todays* [today's] society.

(6)     Being surrounded by other criminals with negative *attitude* [attitudes] will *effect* [affect] your mentality.

(7)     They have to find out a way to do rehabilitation that works and have a positive *affect* [effect] on people.

Almost all occurrences of *loose* in the NICLE material represent the verb and should have been spelled *lose*.

## 8.8.2 Compounds

Whereas Norwegian compounds are regularly spelt as single words, there is a great deal of variation in English. Compare, for example: *school age, schoolboy, school friend, schoolhouse*; *word blindness, word-play, word processor, wordsmith*; *workbook, work experience, workload*. It is difficult to formulate hard-and-fast rules. Dictionaries give some guidance, but this is only helpful up to a point, as new compounds appear and dictionaries cannot therefore contain a full list.

In the light of the NICLE material it is clear that Norwegian students commonly write compounds as single words (or with a hyphen) where this is unlikely or less likely in native English. Examples:

> *crewmembers, deathpenalties, dreamvisions, drugproblems, dutywork, exam-results, feministgroup, healthproblems, knittingneedles, lifequality, livingconditions, livingroom, lunchhours, newscrew, powerstructure, propertyowners, radiochannels, rolemodel, slaveowners, spendingmoney, startingpoint, Sunday-dinner, washingpowder*

Here the problem is a matter of correct spacing. In other cases, there are additional problems, as in:

- *children-movies* (correct: *children's movies*), *dataprogrammes* (correct: *computer programs*), *sivilengineer* (correct: *civil engineer*)
- *prisonpunishment, selfindependent*

The last two forms, which probably reflect Norwegian *fengselsstraff* and *selvstendig*, are unlikely as words in English, regardless of the way they are written.

Problems with compounds thus extend beyond spacing. Compound formation in English and Norwegian, including problems for Norwegian learners, clearly deserves a more detailed study.

### 8.8.3 Use of the apostrophe

Errors in the use of the apostrophe are very common. The problem in distinguishing between *its* and *it's* was pointed out in **8.8.1**. Most often the apostrophe is omitted with *s*-genitives. Some examples can be found in the examples above, e.g. (20) and (44) in **8.7.2**. The apostrophe may also be misplaced, as in:

(8)     Too many of us tend to think that it is somebody ***elses'*** [else's] *responsability* [responsibility].

(9)     To replace religion with television in ***Marxs'*** [Marx's] statement is certainly possible and the statement would still be meaningful, but the statement would be somewhat changed.

(10)     The US Declaration of Independence did not change the slavery or ***peoples'*** [people's] opinions or views when it came to equality.

In (11) below the apostrophe is omitted in the first instances, and misplaced in the second. In the second case the *of*-construction seems preferable. The same is true of (12), where *today's* would be the correct form if we keep the genitive. Note also the incorrect plural form of *life*.

(11)     The opinions of the ***days*** [day's] first meal are without doubt divided. Every person sometime or *another* [other] has had breakfast. Some people think that it is ***the days' most important meal*** [the most important meal of the day].

(12)     How important is television in ***Todays' people's lifes*** [the lives of people today]?

Occasionally the apostrophe is inserted in ordinary plural forms:

(13)     Obviously there's lots of pressure on ***employee's*** [employees] to always improve their effort, but what if they managed to leave work behind by the end of the working day, and more importantly during their ***day's*** [days] off.

(14)     I must emphazise that in my opinion ***such criminals'*** [such criminals] deserve to spend a long time in prison.

(15)     The ***one's*** [ones] who are used to *sit* [sitting] around in front of a computer and read a lot, should get a chance to socialize with others.

The use of the apostrophe to mark the plural is only correct in special contexts, such as: letters (*mind your P's and Q's*), abbreviations (*PhD's* or *PhDs*), and optionally with years (*1990's, 1990s*). Note, incidentally, that *used to* in (15) represents the adjective which is followed by an *ing*-form; cf. **4.2.5**.

## *9. Characteristics of learner language: syntax and discourse*

### 9.1 Grammatical errors

There are many grammatical errors in the NICLE material, both interlingual and intralingual. Some common error types will be briefly surveyed below. The same principles will be used in presenting examples as in Chapter 8 (see **8.1**).

### 9.1.1 Nouns and noun phrases

In examples like those below, definite noun phrases are often used where an indefinite form is required:

(1)     People think ***the nature*** [nature] will protect and save by its own power.

(2)     I think ***the religions role*** [the role of religion] as consolidator was the main reason why Marx maintained that [it] was opium to the masses. ***The religion*** [Religion] comforted the masses by promising a better life after death.

(3)     Even though they are experienced criminals I think that ***the society*** [society] in the long run is going to benefit from helping the criminals, instead of *make* [making] them feel *not welcome* [unwelcome] in ***the society*** [society].

(4)     I believe that we have become too dependent on ***the modern technology*** [modern technology].

(5)     In *the today's society* [the society of today] ***the technology*** [technology] has, in many ways, taken over.

With generic reference of uncountable nouns, English uses the zero article, while the definite form is the rule in Norwegian; cf. Hasselgård *et al*. (2004: 115). In other words, these errors are interlingual. Where there is a postmodifying *of*-phrase, as in the first correction in (5), the definite form is the rule in English as well.
    Another common error type is found with uncountable nouns, as in:

(6)     Personally, I believe that without dreams and ***imaginations*** [imagination], we would be nothing more than robots.

(7)     Eager scientists have been doing scientific ***researches*** [research] on the human brain *since* [ever since] they found out that we actually have brains.

(8)     There ***are researches*** [is research] which *show* [shows] that prisons do fulfill their task *in rehabilitate* [of rehabilitating] criminals.

(9)     Here we need ***a perfect teamwork*** [perfect teamwork], otherwise the treatment will never give results.

(10)    No doubt, students would *win* [gain] a lot through ***a closer cooperation*** [closer cooperation] with both *co-students* [fellow students] and lecturers.

(11)    What they believe is important news, ***are the news*** [is the news] we can read about the next day.

As for English-Norwegian differences in this area, see Hasselgård *et al*. (2004: 90f.). In (11) *news* is treated as uncountable in the first case and in the second as a plural noun.
        More sporadically we find morphological errors of this kind:

(12)    Even natural ***phenomenons*** [phenomena] *and such* [and the like] were mysterious.

(13)    Meeting different kinds of people is a very important part of our ***lifes*** [lives].

Another noun where there is a morphological problem is *stimulus*, where singular and plural forms are used incorrectly (correct: singular *stimulus*, plural *stimuli*).

## 9.1.2 Pronouns

Errors in the use of reflexive pronouns are fairly common. Some of these are purely morphological, as in:

(14)    As a boy he made a pact with ***himselves*** [himself] that he would become the greatest artist the world had ever seen.

(15)    It may say something about the way we perceive ***ourself*** [ourselves].

See the list of correct forms in Hasselgård *et al*. (2004: 126). In the following example there are multiple problems:

(16)    It is really up to the students themselves to make a schedule that gives *you* [them] the theory *you* [they] need and ***get theirself*** [also to get] experience.

One problem here is the unnecessary switch between third- and second-person pronouns; it is better to stick to the third person, which is used at the beginning of the sentence. The third-person reflexive pronoun is correct in the first instance. In the second instance, the form is incorrect, and the reflexive form is best omitted. By inserting *also* plus the infinitive marker *to* we get a somewhat clearer syntactic structure.
        Other types of errors involving reflexive pronouns are illustrated in:

(17)    Society cannot handle it, or maybe even ***themselves*** [they themselves] cannot handle it.

(18)    The opinion was that ***themselves*** [they themselves] and these beings were either living simultaneously on earth, or the creatures had been there before them.

There is also a tendency to use reflexive verbs where non-reflexive verbs are required, as in:

(19)    Therefore we do not ***worry ourselves with*** [worry about] the things we do not know.

(20)    The recruits in the US army ***enlist themselves*** [enlist] for a period of 2-20 years.

(21)    The students should probably ***specialise themselves*** [specialise] in a few subjects instead of learning a little bit of this and a little bit of that.

(22)     There has been a shift in *focuse* [focus] away from women and on to gender, which ***concerns itself*** [is concerned] with man and woman in mutual interdependence.

(23)     Criminals need help in order to re-adapt to society, there is no point in locking them up for years and years, hoping everything will ***fix itself*** [sort itself out].

These errors are interlingual. Note that Norwegian versions would contain reflexive verbs: *uroe seg, la seg verve, spesialisere seg*, *dreie seg om, ordne seg*. In (23) we can keep the reflexive form with a change of verb. For English-Norwegian differences in the use of reflexive verbs, see Hasselgård *et al*. (2004: 131f.).
        Another error type which occurs repeatedly in the NICLE material is found in relative clauses. Examples:

(23)     There are not many people who will hire a person***, which*** [who] has done something bad like *steeling* [stealing].

(24)     Maybe the soldiers ***which*** [who] usually are left at end of *an* [a] year *gets* [get] better training.

(25)     For many, for example drug addicts***, which*** [who] do not have a roof over their heads every day, this must be like a vacation at a *three stars* [three-star] hotel?

(26)     One of the differences from "the good old days" ***who*** [which, that] *attracts* [attract] people is that *the television* [television] *is showing* [shows] the *watchers* [viewers] what the generations before the 20th century had to imagine.

(27)     If Marx was alive today I think he would say that the human race *have* [has] once again invented something to control other people, this time a box ***who*** [which, that] *are telling* [tells] you what to do and how to be.

We find both *which* used with personal nouns as in (23) to (25) and *who* combining with personal nouns as in (26) and (27). For a description of the use of the relative pronouns, see Hasselgård *et al*. (2004: 339ff.).
        Note, incidentally, that a comma may be inserted incorrectly before restrictive relative clauses, as in:

(28)     They were fooled into a revolution***, which*** [which] they knew nothing about.

(29)     This is a view of university education***, that*** [that] many people, *righteous* [rightly] or not, tend to have.

Another clear case is found in (23) above. These relative clauses are clearly restrictive (cf. Hasselgård *et al*. 2004: 342), and the comma should be dispensed with.

### 9.1.3 Verbs and verb phrases

A common type of error in verb phrases is the overuse of the progressive form, as in:

(30)     People in general ***are being*** [are] aware of their need for peace and have probably had more than enough of *the city life* [city life], which is *stressing* [stressful] and noisy.

(31)     A person freely *enrols himself* [enrols] for military service whereas with the Norwegian system he ***is being*** [is] conscripted.

(32)   Compared to previous times we ***are almost behaving*** [almost behave] like robots sometimes.

See also examples (26) and (27) above. As for the overuse of the progressive, see Hasselgård *et al.* (2004: 184) and the paper by Johansson and Stavestrand (1987).

Modal auxiliaries are also problematic, particularly *shall* and *should*, which differ greatly from their Norwegian cognates *skal* and *skulle*; see the description of the uses of *shall* and *should* in Hasselgård *et al.* (2004: 199ff.). In the NICLE material, errors are found particularly with *shall*, as in:

(33)   When you commit an offence and the Court decides that you ***shall*** [must] serve a prison sentence, you ***shall be dreading*** [will dread] it.

(34)   There should be bad food served in the kitchens, you ***shall*** [should] *make* [prepare] it yourself and you should only have a few basic ingredients.

(35)   People should dread going to prison, *ofcourse* [of course] no physical or mental harm ***shall*** [should] *befall* [be done to] the inmates but *it* [prison] should be an environment of solitude.

As for Norwegian *skal* and its correspondences in English, see further **4.3.2**.

Another area which frequently causes problems is verb complementation. Examples have already been given in connection with the discussion of colligation errors; see **8.1.6**. These are some more examples:

(36)   Who can ***admit to be*** [to being] such a person?

(37)   Hopefully the fear of being imprisoned will ***prevent*** possible criminals ***to break*** [from breaking] the law.

(38)   Does it ***deprive*** us ***from*** [of] social contact?

In (36) *admit* requires the preposition *to* plus an *ing*-form, in (37) *prevent* must be followed by the preposition *from* plus an *ing*-form, and in (38) *deprive* requires the preposition *of* rather than *from*.

As with nouns, we sometimes find morphological problems, e.g.: *lead* rather than *led* used for the past-tense and past-participle form of *lead*; *payed* rather than *paid* used for the past-tense and past-participle form of *pay*; *broadcast* treated as a regular verb.

### 9.1.4 Adjectives and adverbs

In Norwegian there is no morphological counterpart of the English distinction between adjectives and adverbs (ending in *-ly*). Not surprisingly, the distinction turns out to be problematic for many learners. Most often the *-ly* ending is omitted, as in:

(39)   In the event of a future war situation, the technology will enable an opponent to act fast and ***efficient*** [efficiently].

(40)   The authorities often *made* [took] advantage of this trust *to* [in] God and used the gospel ***manipulative*** [manipulatively] to keep *inhabitants* [people] calm.

(41)   We can see it even more ***obvious*** [clearly] in our modern society.

(42)     I really think that the minds of kids today are overstimulated. When can they ***possible*** [possibly] have the time to try and activate themselves?

(43)     ***Preferable*** [preferably] one should be effective and know a lot about 'down-to-earth' things.

(44)     The Norwegian army *don't* [doesn't] train their soldiers ***good*** [well] enough to meet the challenges they are faced with on the battlefield.

(45)     When people are weary, disagreements and quarrels may ***easier arise*** [arise more easily].

These are clearly contexts where adverbs are called for; cf. Hasselgård *et al.* (2004: 237ff.).
Occasionally we find may find an *-ly* form used 'hypercorrectly' in a context where an adjective is required:

(46)     One may refuse to serve in the ***traditionally*** [traditional] way (and do some kind of a community service instead).

Problems with adjectives are illustrated in:

(47)     If your mind is forced to work too hard during the most of your ***awake*** [waking] hours, there will be little energy left for dreaming.

(48)     If someone got killed in the Norwegian Saga period, ***the dead's family*** [the victim's family, the family of the deceased] had the right to kill anyone in the killer's family.

*Awake* is one of the adjectives which do not occur in attributive position (cf. Hasselgård *et al.* 2004: 225). In (48) we find a nominalised adjective used in the Norwegian manner; the use of nominalised adjectives is far more restricted in English (cf. Hasselgård *et al.* 2004: 228ff.).
Many problems with adverbs have to do with where they go in the sentence; see **9.1.9**.

## 9.1.5 Concord

Perhaps the most common error type of all is lack of concord between subject and verb. Some errors appear to be mere slips which could easily be corrected by the learner. Often there is, however, some complication. The noun governing the concord relation may be separated from the verb, as in:

(49)     ***One*** of the reasons for the bad results ***were*** [was] the children's *TV-habits* [TV habits].

(50)     In the most civilised part of the world it looks like our ever growing ***knowledge*** about our surroundings ***are*** [is] turning people away from religion.

(51)     The chain continues with ***effects*** on the weather which ***becomes*** [become] more marked.

(52)     They do not adopt the common ***norms and rules that is*** [are] required to be a part of the community.

In (49) the concord relation is determined by *one*, but problems are caused by the intervening plural forms; (50) is a similar example. In the next two examples the concord relation is determined by the antecedent of the relative pronouns. The example below contains three concord errors; the last two are of the same kind as in (51) and (52).

(53)    In a world where **rationality and technology plays** [play] a major role, **dreams**, which often **consists of** [are full of] symbols, **plays** [play] an equal part.

Another problem is caused by the coordinated noun phrase in subject position; since *rationality* and *technology* denote different things, the plural seems more natural here.

Often concord problems are caused by the type of noun in subject position. Some examples are:

(54)    The ideal **family female** [housewife: cf. familiekvinne] **have** [has] *been killed* [become extinct].

(55)    Not **everybody manage** [manages] *such a strain it is* [the strain it is] for both parties to *being* [to be] married, having children or getting *threw* [through] disease or crisis.

(56)    The question is a really hard one, because **there are not really anyone** [there really is not anyone] that can say they agree only with one side.

(57)    **People** all over the world **is** [are] working only parts of the day, and then they can do whatever they want.

(58)    In contemporary society, **people seems** [seem] tired of the whole subject matter of female versus male.

(59)    **Police is** [the police are] the number one physical crime- stopper with millions of police officers worldwide.

(60)    In Norway **the police investigates** [the police investigate] all reported criminal actions.

In the first three examples there may be a pull towards the plural because the subject noun phrases may refer to groups of people. But the concord relation is determined by the singular form. In (57) to (60) we find the opposite type of error. Although they lack a plural ending, *people* and *police* are treated as plural nouns. Note that we can easily insert a numeral: *fifty people, twenty police*.

For a survey of subject-verb concord in English, including some of the problems touched on above, see Hasselgård *et al.* (2004: 265ff.). To conclude this section, let's look at some of the findings in a doctoral thesis on learner problems with concord in English (Thagg 1985). The thesis deals both with errors in subject-verb concord and with some minor categories of concord problems. Only the former will be considered here. Although the thesis was written more than two decades ago and was based on material produced by Swedish students of English, most of the findings are no doubt applicable more generally.

Ulla Thagg examined concord errors from spoken and written English, supplemented by elicitation tests given to both Swedish learners and native speakers of English. The spoken material was taken from the retelling of a story and the description of pictures, altogether 15 hours of recorded speech. The written material was gathered from translation tests and compositions. The aim of the elicitation tests was to provide more information on problematic areas. The native speakers were also asked to assess the gravity of errors selected from the learner material.

The highest error density was found in spoken production and the lowest in the translations, presumably because the students could focus more on form in the latter case. The speakers, on the other hand, had little time for planning and had to focus on content rather than formal accuracy, which led to many slips (performance errors).

The learners and the native speakers were found to have some problems in common: the dilemma posed by grammatical vs. notional concord, attraction exerted by other nouns than the head of the subject noun phrase, distance between the agreeing elements, and areas of variable concord. By way of conclusion, Thagg distinguishes between the errors which are native-like and those which are characteristic of the learners; see Figure 9.1

LEARNERS' CONCORD BEHAVIOUR

Native-like        Non-nativelike

Non-contrastive       Contrastive

He was one of those enviable persons who *seems* to be gifted with a natural balance of mind.
One gets rid of the ordinary vacation public, who *does* not bathe.
I think travelling abroad and the close contact with foreign people *develops* one's character.
A series of stamps *are* to be sold at a decent price.

Charles *think*
He *give*
The boys *looks*
They *uses*

Everybody *work* together
The money which *were* needed

Deemphasize in teaching
Do not penalize in tests

Emphasize in teaching
Penalize in tests

Figure 9.1 Native-like and non-nativelike concord behaviour (based on Thagg 1985: 191)

Nativelike behaviour includes the use of the singular in relative clauses following *one of* plus plural noun and the choice of verb form with collective nouns, coordinated noun phrases, and subject noun phrases of the type *a series of* plus noun. Non-nativelike errors can be non-contrastive or contrastive, i.e. in the latter case connected with differences between a form in the mother tongue and the corresponding form in the foreign language (e.g. *everybody*, cf. *alle*; *money*, cf. *penger*). In general, Thagg found that there are conflicts between grammar/textbook norms and actual language use among native speakers of English.

## 9.1.6 Dummy-subject constructions

Whereas Norwegian has the dummy subject *det*, English makes use of the two forms *it* and *there*. These are often confused by Norwegian learners; see Hasselgård *et al.* (2004: 319ff.). Note that the problems extend beyond the choice between *it* and *there*; cf. the discussion of

correspondences of *det* in **5.7**. In the NICLE material we find examples where dummy *it* is used incorrectly:

(61)   ***It still has not passed a hundred years*** [Less than a hundred years have gone by] since the women in my, and most other European countries, were allowed to use their democratic right to vote.

(62)   There are not many people who will hire a person *which* [who] has done something bad like *steeling* [stealing]. ***It is*** [there is] always a risk that it can happen again.

In (61) we need to restructure the sentence altogether and choose a different subject, in (62) we find a clear case of dummy *it* used instead of existential *there*. The opposite type of mistake is found in:

(63)   I think ***there*** [it] will take more than a decade or a couple of governments to figure out what works.

(64)   ***There are not everybody who are*** [Not everybody is] capable of nursing a mentally or physically (or both) handicapped child.

In (63) the time expression requires *it* rather than *there*, in (64) we must restructure the sentence.

### 9.1.7 Preposition plus complement

According to an old pedagogical rule, "preposisjon pluss *at* er som hund og katt, preposisjon pluss *å* kan heller ikke gå", or more precisely, the corresponding forms in English. Whereas a preposition in Norwegian can be followed by both infinitives and *at*-clauses, these kinds of structures are ruled out in English (with very few exceptions). The difference may lead to errors, as in:

(65)   The question is *if* [whether] the prison system is outdated and *if* [whether] there are other ways ***of rehabilitate*** [of rehabilitating] criminals.

(66)   I think that *the television* [television] is a very good tool for manipulating the masses, and the pictures from the 11th September crash in New York going over the *TV-screens* [TV screens] all over the world *was* [were] the *terrorists* [terrorists'] way of manipulating us, making us scared and at the same time ***aware of that*** [aware that] there *exists* [is] danger that could *harm* [threaten] everybody.

(67)   But even if we ***agreed on that*** [agreed that] *the university degrees* [university degrees] are theoretical, it does not have to imply that they are of little value.

(68)   There are reasons ***for and against that*** [for and against the claim that] crime does not pay.

(69)   There *has* [have] also been formed many *extremist-groups* [extremist groups], and they don't care ***about that there actually is*** [that/whether there actually is, about there actually being] a law in that area too.

In (65) the infinitive must be replaced by an *ing*-form. The other examples illustrate the incorrect use of preposition plus *that*-clause. In (66) and (67) we can simply drop the

preposition, an option which is not possible in (68); here we can instead insert a noun (68). In (69) we can drop the preposition or choose an *ing*-construction.

For comments on differences in this area between English and Norwegian, see **6.3** and Hasselgård *et al.* (2004: 349).

### 9.1.8 Indirect questions and comparative clauses

Some errors in the NICLE material appear to reflect the use of a complementiser in Norwegian. These include indirect questions, as in:

(70)   Have you ever thought *of* [about] ***how many people who're working*** [how many people are working] in the *data* [computing, computer] business?

(71)   Before, people did not know how big the earth is or ***how many different peoples that exist*** [how many peoples exist] in the world.

Cf. the use of *som* in Norwegian: *hvor mange mennesker **som** arbeider*, *hvor mange forskjellige folkeslag **som** eksisterer*.

In comparative clauses introduced by *than*, learners often insert *what* reflecting Norwegian *hva*:

(72)   We can always try to make it closer to perfect ***than what*** [than] it is.

(73)   There are NOT *much* [many] more *drug-addicts* [drug addicts] in the Netherlands ***than what*** [than] *there is* [there are] in Norway or the United States.

(74)   The prison standards are often considerably higher ***than what*** [than] they came from.

(75)   Thanks to the technical revolution the crimes today are more advanced and much more difficult to discover ***than what*** [than] they were back in the old days.

(76)   We will always need to wind down and think about other things ***than what*** [than] we are working with.

(77)   Most professional criminals put down more work in what they do ***than what*** [than] most law-abiding citizens will ever do.

(78)   Altogether, the resources and the big money would be in much more needy places ***than what*** [than] seems to be the case today.

As this error type is recurrent and does not seem to have been commented on in contrastive grammars, it is a topic that deserves a more detailed study.

### 9.1.9 Word order

Considering that there are major differences in word order patterns between English and Norwegian (see Hasselgård *et al*. 2004: 298ff.), it is not surprising that we find word order errors in the writing of Norwegian students of English. These are examples of subject-verb inversion used in the Norwegian manner:

(79)   ***Perhaps was the successful snatch*** [The successful snatch was perhaps; cf. kanskje var …] what triggered a more serious crime, changed their lives and made them suffer.

(80)   Of course ***should they*** [they should; cf. selvfølgelig skulle de] no longer stay at home.

(81)   Here *is probably an abortion* [abortion is probably] a good solution.

(82)   Is it because we are satisfied with the way things are, or is it because we long for something more, dream of something bigger and better? *In my point of view is it the latter* [In my view it is the latter].

(83)   *Even in schools do children learn* [Even at school children learn] to use computers from the age of 6, and as an adult you are a "loser" if you cannot manage to use a PC.

Subject-verb inversion is far more restricted in English than in Norwegian, but it is required in the following example after the opening restrictive element:

(84)   *Only in Physical Education we do have* [Only in Physical Education do we have] a practice exam where we are suppose to teach other students.

Here the learner has overgeneralised the typical pattern of declarative sentences in English. In other words, this error is intralingual.

Although there are some errors in the placement of subject and verb, the most common type of word order problem has to do with adverbial placement. Though the order is not completely unacceptable, it would be more natural to place the adverbial somewhere else in:

(85)   *I will in this essay focus* [In this essay I will focus] on my degree*.*

(86)   *An army will from now on* [From now on an army will] be defined as any large number of fighting men, their equipment and a system of leadership.

(87)   *They are on the contrary* [On the contrary they are] unwanted elements that society spends a great amount of time and effort to eliminate.

(88)   Strong words about education, but even so I must say that I *to some degree agree* [agree to some extent].

(89)   We don't *to the same extent as before make images in our minds ourselves* [make images in our minds ourselves to the same extent as before]. When we read a book it is our imagination that *to a strong degree creates the setting* [determines the setting to a great extent].

(90)   *The military service is in todays Europe practiced by many nations* [Many nations in today's Europe have a system of military service].

According to Hasselgård *et al*. (2004: 292), adverbials realised as phrases are best placed in initial or final position. In (85) and (86), where the reference is to something taken as given (the essay, the stage of writing), the adverbial goes well in initial position. Initial placement is also natural with the linking adverbial in (87). In contrast, the degree adverbials in (88) and (89) are best moved to final position. In (90) we need to rewrite the whole sentence.

In her study of thematic choice in a selection of NICLE essays and in native-speaker texts, Hilde Hasselgård (forthcoming) makes the point that initial adverbials are found more commonly in the learner material. The initial adverbials in the example below are not wrong, but the suggested reordering makes the text clearer:

(91) ***In the beginning of the fight for women*** it is clear that the *feminist's* [feminists'] work resulted in positive achievements. But ***during the last years*** it is perhaps correct to say that, due to their impatience to succeed, they have done more harm to the cause of women than good.

This is a suggested rewording:

(91a) It is clear that, initially, the feminists' work resulted in positive achievements. It is, however, perhaps correct to say that in the last few years, due to their impatience to succeed, they have done more harm to the cause of women than good.

The placement of adverbials is, however, a complex matter where it is not possible to draw a clear borderline between what is right and wrong, and where it is important to take the wider context into account.

## 9.2 Information structure

As explained in the discussion of word order in Hasselgård *et al*. (2004: 298ff.), there are some important principles which guide the placement of words in sentences: in addition to syntactic principles, in particular the information principle and the principle of end weight. In the NICLE material we find examples which go against one or both of these principles:

(92) Similarly to religion, television *has clearly* [clearly has] an element of entertainment, that is for sure. It is possible to forget about family and work problems while watching movies, news or sport. An unhappy friend or a lonely grandmother *is also to be forgotten easily* [is easily forgotten]. In the same way, ***getting less politically involved may be an effect of watching much television*** [another effect of watching too much television is that one may get less politically involved].

(93) ***To which extent prison it self has any preventative effect has also been discussed*** [It has also been debated to what extent prison itself has any preventative effect].

(94) Within a cultural group, and between cultural groups, you will find very different opinions on these matters. ***The fact that these are "hot issues" seems to be agreed upon*** [There seems to be agreement that these are "hot issues"].

(95) First of all, ***the leaving of small children in front of a television screen without some kind of supervision is terrible*** [it is terrible to leave small children in front of a television without some kind of supervision].

(96) ***To treat those who already are experienced in the field of crime; it takes more work to get them to start a different life*** [It takes more work to treat those who are already experienced in the field of crime and get them to start a different life]. Maybe *it* [there] is more to it than just *to make* [making] them *to quit* [quit] criminal activities.

As the preceding context in (92) is about effects of watching television, it is better to reorder (and reword) the elements in the last sentence. Note also that the placement of *clearly* is unfortunate; adverbials in English are not normally placed between a verb and its object. The other examples contain long and complex subjects, which are best postponed (with the insertion of a dummy subject), in agreement with the principle of end weight.

Word order and information structure are important factors contributing to the well-formedness of a text. Word order and related syntactic means should not be used in a way

which may confuse the reader. In **5.6** attention was drawn to the more common use of cleft constructions (*it*-clefts and *wh*-clefts) in Norwegian and Swedish than in English. Therefore it comes as no surprise that learners overuse cleft constructions in English. In a paper by Boström Aronsson (2003) it is shown that the unmotivated use of a cleft construction may have unwanted effects.

## 9.3 Cohesion

A distinction is commonly made between ***coherence*** and ***cohesion***, where the former applies to texts which are connected in a clear and understandable way and the latter refers to formal markers which contribute to the coherence of a text; see Hasselgård *et al*. (2004: 400). Here attention will be briefly drawn to two important categories of formal markers: pronouns and connectors.

### 9.3.1 Pronoun references

For a text to be coherent it is crucial that the use of pronouns is clear and consistent. Pronoun use is often unclear in the NICLE material, as in:

(97)    You remember that you forgot to brush your teeth after breakfast so you go into a supermarket to buy ***one*** [a toothbrush].

(98)    It seems to me that at the moment crimes pay, if you get away with ***it*** [them].

(99)    Adults worked full days and evenings in the old days too, but they could not place their children in front of the television or video while doing ***it*** [their work].

(100)   The trains carry the same *amount* [number] of people but if 1/3 of these do not pay, it means that ***they*** [those running the trains] do not get the money ***they*** are supposed to and can therefore not give the service to ***their*** customers, as ***they*** would like.

(101)   If ***we*** look at it like this, ***one*** [we] may see that even in the technological and industrial field dreams help to create new technology.

In (97) we understand that the reference must be to a toothbrush. Such 'sloppy' references are common in conversation, where possible misunderstandings can easily be cleared up, but they are best avoided in writing, where there is no immediate contact between the writer and the reader. Inserting *them* in (98) makes the reference clearer; here *it* suggests that the writer was thinking of the process of committing crimes. *It* in (99) lacks a co-referent form and is best replaced by a full noun phrase. In (100) *they/their* must refer to those responsible for running the trains, rather than the trains, which is the interpretation suggested by the wording. In (100) there is an unmotivated change of subject. The next example is more complex:

(102)   If ***you are*** [a person is] caught it usually means that ***a person*** [he/she] will get a conviction, and be imprisoned. This means that ***this person*** [he/she] will *loose* [lose] ***its*** [his/her] freedom, and ***this conviction*** will follow him/her the rest of *his/hers* [his/her] life.

In the first place, there is an unmotivated switch from *you* to *a person*. It would be preferable to keep the same type of reference throughout. As the forms *he/she* and *his/her* are not very elegant, it would be even better to use a plural noun. A further problem is the distance between *the conviction* and the first mention of *conviction*. At this point we rather expect a

reference to the whole experience (conviction plus imprisonment). After adjustments, the text will run as follows:

(102a) If people are caught it usually means that they will get a conviction, and be imprisoned. This means that they will lose their freedom, and the experience will follow them throughout the rest of their lives.

Some students have a tendency to overuse *this*, as in the following passage, where there a number of other problems as well:

(103) Looking at *abortions* [abortion] from an ethical point of view, it's clearly wrong. Killing is wrong in any form, we are not allowed to do it. *It's* [There's] also a clear contrast to *the one's* [those] who can't have children. They spend so much money and time trying to get pregnant, when on the other *side* [hand] there's a problem *and* [in that] we are taking lives.
    We have to be careful concerning *abortions* [abortion] because it might lead to a general attitude among people that if something happens we can always have an abortion. ***This*** is a very *dangerrous* [dangerous] attitude to have among young people, many will then just sleep around and ***this*** again will lead to venereal diseases. ***This*** is something to be concerned *with* [about].

Notice, first of all, that there is a problem with the relationship between *abortions* and the intended co-referent pronoun *it*. As the text has to do with the phenomenon in general, it is preferable to use the singular form *abortion*, which makes the use of *it* correct. The pronoun *this* is found three times in just a few lines. It is a very convenient form to use when we want to refer to what has been said earlier in the text, but what does the last instance of *this* refer to here? Here is a suggested rewording of the passage:

(103a) From an ethical point of view, abortion is clearly wrong. Killing is wrong in any form, and we are not allowed to do it. There's also a clear contrast in relation to those who can't have children. They spend so much money and time trying to get pregnant. At the same time there's a problem in that we are taking lives.
    We have to be careful concerning abortion because it might lead to a general attitude among people that if something happens they can always have an abortion. This is a very dangerous attitude for young people to have, and many will then just sleep around, which in turn will lead to venereal diseases. These are problems which we should be concerned about.

After the rewording, only one instance of *this* remains. In addition to addressing the points which have already been made, the new version solves some other problems, including these: the first sentence is simplified, eliminating the dangling participle (see below) and at the same time resolving the reference problem; two cases of comma splices (see **9.7**) are removed.
    ***Dangling participles***, or unattached participle clauses, may cause uncertainty with respect to the understood subject or may even result in an unintended meaning; cf. Hasselgård *et al*. (2004: 367f.). Although they are not uncommonly found in native-speaker writing, they are best avoided by foreign learners. These are some examples from the NICLE material:

(104) Maybe there should have been a requirement before ***going*** [we go] to a *teaching* [teacher training] college to have a few years in between high school and college?

(105)   ***Without seeing*** [If you do not see] both sides of a situation, it is impossible to make a just decision on what you believe is right.

(106)   There are some questions that need to be answered before ***deciding*** [we decide] if it's right or wrong.

(107)   Life was harder, but it must have been a golden age for imagination and dreaming. Just think of all the wonderful thoughts that could run through a person's head while ***doing*** [he/she was doing] some work.

(108)   Just by ***being*** conscious of the important task we have in taking care of our earth, ***a lot can be done*** [we can do a lot].

As we see here, we can get rid of dangling participles by inserting a subject pronoun for the verb in the participle form. Alternatively, the main clause can be rephrased as in (107), so that the subject agrees with the understood subject of the *ing*-form. The problem with dangling participles is thus that not that pronouns are misused, but that they are omitted where they are required (or preferred).

### 9.3.2 Connectors

In a paper on linking adverbials (adverbial connectors) Altenberg and Tapper (1998) report on a study comparing essays in the Swedish learner corpus SWICLE and the native-speaker reference corpus LOCNESS. They found some differences in the use of particular connector types, but the main conclusion was that the Swedish learners lacked register awareness and had a tendency to mix formal and informal elements. Building on their study, Anne-Berit Eia (2006) explored the use of linking adverbials by Norwegian learners on the basis of material from NICLE and LOCNESS.

Like Altenberg and Tapper, Eia found some differences between individual adverbials. The Norwegian students overused less formal linking adverbials (*for instance, on the other hand, still, of course, so*) and underused more formal linking adverbials (*for example, i.e., such as, however, yet, also, firstly, therefore*). In general, they tended to adopt a more informal writing style than the native students.

Apart from linking adverbials, Eia examined the use of the conjunctions *and* and *but* in initial position. She found that the two conjunctions made up about a third of all cases of initial connectors (including both linking adverbials and *and/but*), as against about 10% for the native speakers, and she hypothesised that *and/but* may be chosen sometimes by the learners rather than linking adverbials. Although *and/but* do occur in initial position also in the writing of native speakers, the overuse among the learners is another element contributing to a more informal style. These are some examples found in an essay discussing the value of university degrees:

> ***Most university degrees are theoretical and do not prepare students for the real world. They are therefore of very little value*** (topic)
> This is a view of university education that many people, *righteous* [rightly] or not, tend to have. *And* most students have more than once questioned the value of sitting passively in the auditorium, … .
> *But* the fact is luckily more nuanced than this … .
> *But* when it come to questioning the value of a university degree … . *And* this will normally be tried out … . *And* this is not given … .
> There are for sure … .

Clearly changes for the better could be introduced, to prepare students for their careers. *But* a crucial point has to be made … . *But* even so, the need *of* [for] flexibility does not require hours of one-way conversation in an auditorium as its method.

Four of the five paragraphs contain sentences with initial *and/but*. The second and third paragraphs open with *but*, and the last paragraph includes two instances of sentence-initial *but*, including one in the final sentence of the essay. Note, incidentally, that *this* in the first sentence, referring to the essay title, is not a good opening; we expect a self-contained text to give an introduction to the topic. The ending is also weak and could be rephrased in this way:

> *But* even so, the need for flexibility does not require a method which demands hours of one-way conversation in an auditorium.

The rephrasing allows the focus to fall on the main point of the essay.

## 9.4 Writer/reader visibility

In dealing with word frequencies in Chapter 8 (**8.4**) we found that first- and second-person pronouns tend to be overused by learners. Special attention was drawn to the overuse of the sequence *I think*. We can regard this as reflecting a more general tendency in essays by language learners, referred to in a paper by Stephanie Petch-Tyson (1998) as writer/reader visibility. Petch-Tyson compared four ICLE corpora and the US subcorpus of LOCNESS with respect to a number of features of writer/reader visibility, including:

> First- and second-person reference
> Reference to the writer's mental processes (*think, believe*, etc.)
> Monitoring of information flow (*you know, I mean*, etc.)
> Emphatic particles (*just, really*)
> Fuzziness (*and so on, etc.*)
> Reference to the situation of reading/writing (*this X, here, now*, etc.)
> Imperatives
> Questions

These features are overt markers indicating the presence of the writer/reader in the text. Petch-Tyson found that all of these features were more common in the learner corpora than in the native-speaker reference corpus. Typical occurrences of *I* plus verb referred to the writer as organiser of the text. Petch-Tyson concluded that "the learner writers are much more overtly present within the discourse than the NS [native speaker] writers".

Petch-Tyson's study did not include NICLE, but her findings are no doubt applicable also to the essays by the Norwegian learners. Figure 9.2 below gives part of a concordance of *I will* in NICLE (there were 112 occurrences in all). Apart from a handful of occurrences, these are all concerned with the writer operating within and planning the text.

| N | Concordance: *I will* |
|---|---|
| 1 | on our minds. And **as an additional point I will add** the extreme focus on appeara |
| 2 | on ? **There are moral and ethical aspects I will adress** in order to answer these |
| 3 | regarding marriage and lifestyle. **I will also add some of my personal opi** |
| 4 | ements - incorporated in the human mind. **I will also illustrate** it with an examp |
| 5 | idually and together with other persons. **I will also point out** that a profession |
| 6 | on in the future. **In conclusion, the way I will answer my question** in the first |
| 7 | should rehabilitate them.' **In this essay I will argue** for and against this claim |
| 8 | re it to avoid the problem of pollution? **I will base my essay** upon Victor Hugo's |

| | |
|---|---|
| 9 | emember grandfathers last story, so that I will be able to tell it to my childre |
| 10 | ilitary service? **These are the questions I will be discussing in this essay**. Sin |
| 11 | eness and imagination. **In the following, I will briefly draw a line** from early h |
| 12 | evision makes me even stronger! Tomorrow I will buy myself a new pair of headset |
| 13 | tion. Well**, the way society works today, I will certainly say** that it has made i |
| 14 | mprisoned, than outside the prison yard. **I will come back to the reason why**, la |
| 15 | d in a totally different way than today. **I will concentrate** on the Scandinavian |
| 16 | his development affects people. **Finally, I will conclude** with some "modest" su |
| 17 | ave passed the thirties for most people. **I will conclude** with that it has never |
| 18 | ing in this essay. Since I am Norwegian, **I will consentrate on** the Norwegian sy |
| 19 | has triggered my imagination. Hopefully I will continue to read books for the r |
| 20 | s movement - the feminists **In this essay I will deal with** the period starting fr |
| 21 | of development of our species. But **first I will define** what I mean by dreaming. |
| 22 | in the real world to learn how to teach. I will definitely not feel that I am su |
| 23 | eriences are bound to shine through when **I will discuss the topic**: "Abortion - r |
| 24 | al creativity and fantasy. **In this essay I will discuss** the fact that some peopl |
| 25 | it? Is it really worth it? **In this essay I will discuss** the advantages and disad |
| 26 | ad of rehabilitating them? **In this essay I will discuss** these questions. The pri |
| 27 | p from society after committing a crime. **I will discuss this matter**, considering |
| 28 | from the rest of society. **In this essay I will discuss** wether the contemporary |
| 29 | ld keep general conscription or not. **Now i will discuss** wheter an army should b |
| 30 | ith this complex problem. **In this essay, I will discuss** whether armies should co |

Figure 9.2 Part of a concordance of *I will* (NICLE)

Another related paper, by Virtanen (1998), deals with the use of questions in some ICLE corpora and in LOCNESS. Virtanen distinguishes between two functional types of questions:

- topical questions, which have a text-organising function;
- rhetorical questions, which express the writer's attitude.

Both types encourage reader involvement. Topical questions are often found in the opening paragraphs of essays. Essay-final questions include both topical and rhetorical questions. Topical questions at the end have a summarising function.

Though there is some internal variation, Virtanen finds that the learners tend to use more questions than the native speakers. She observes that "[a]n overuse of questions can reduce their argumentative value and increase the often more informal style of their writing" (p. 105). No detailed quantitative study has been done on questions in NICLE, but there is no doubt that many students use questions a great deal. These are some examples of questions in different sections of NICLE essays:[8]

(109)   Marx once said that religion was the opium of the masses. If he was alive at the end of the 20th century, he would replace religion with television. ***Why did Marx say that religion was the opium of the masses and to what extent is it possible to replace religion with television?*** (first paragraph)

(110)   Ever since we were introduced to *the television* [television] we have been fascinated by it. ***But what is so special about this device, why has it become such a significant part of our everyday lives? And to what extent has it affected our lives, both positively and negatively?*** (first paragraph)

---

[8] Note, incidentally, that the change of *the television* to *television* in (109) to (111). The form without the article is preferable where the reference is to the general phenomenon. The definite form is appropriate in examples like: *The television is on the table. The television was unplugged. He switched on the television.*

(111) ***So what good can this device do?*** Well, first of all, *the television* [television] really gives you a sense of being part of a culture. Important happenings are being *broadcasted* [broadcast] as they take place. In addition, it can be a great source of information and knowledge that *go* [goes] far *behind* [beyond] the information and knowledge you get from the radio. It is also clearly a *socialisor* [socialiser]; by watching *the television* [television] you can always find things to talk about that you have seen on television. And it can be very social gathering in front of the television. *The television* [Television] is also very useful as a baby-sitter, and easily replaces the bedtime stories by *imposing* [giving] the children a sense of ethics and *moral* [morality]. ***But is there something we are missing?*** (middle of essay)

(112) In the end, it is up to the viewer him/herself to decide what to watch and what to *avoid* [avoid watching] on their *TV-screens* [TV screens]. Our western society prides itself *with* [on] being the land of the free. ***Has this freedom gone to our heads?*** Our entertainment is no longer the innocent game show that the whole family would watch. ***Do we really want a society without borders, without limitations?*** I think not. (last paragraph)

(113) ***So does crime pay off then?*** Obviously not for those who get caught every time. For the rest it certainly appears it does. But *one thing* [there is one thing] you definitely should remember before considering a new career in crime: most professional criminals put down more work in what they do *than what* [than] most law-abiding citizens will ever do. (last paragraph)

In (109) and (110) we find topical questions which the writer sets out to answer in the essay. The next example opens with a question which defines the topic of the paragraph and ends with a question introducing the next topic. In (112) we find rhetorical questions in the final paragraph of an essay, while the question in (113) is an example of the summarising function of questions.

To conclude, there are marked differences between our learner corpora and the native-speaker corpora as regards the frequency of features of writer/reader visibility. Should we regard this aspect of learner language as wrong? This is clearly not the case; features of writer/reader visibility occur among native speakers as well, though they vary depending upon the type of text. What is important is to use the features in an appropriate way and maintain a consistent tone throughout the text.

## 9.5 Stance

Expressions of author stance could be regarded as another feature showing the presence of the writer in the text. In a study of NICLE texts and native-speaker reference texts Hilde Hasselgård (forthcoming) reports that extraposition is overused by Norwegian learners, and she relates this to the discourse functions of extraposition (for an account of extraposition, see Hasselgård *et al.* 2004: 323). This structure is commonly used for the thematisation of evaluation or opinion (*it is hard/difficult/easy to, it is important to/that, it is strange that/how*, etc.), but also for expressing possibility/necessity, truth/fact, subjective evidence, etc. Subjective stance is also commonly expressed by first-person clauses (*I believe, I think, I guess, I would say*, etc.), a typically conversational feature. In addition, there are many other markers of stance (*certainly, surely, of course*, etc.). Hasselgård notes that "the adverbials that are most frequent in conversation are also most clearly overused in NICLE in relation to the other genres" (p. 10). Other voices are also present in the texts, including the addressee (*you could/might say that, you think, you may ask*, etc.), and the addressee is overtly addressed by questions and imperatives. In agreement with the findings of Petch-Tyson (1998), Hasselgård

finds that "the NICLE essays exhibit an interactive writing style with a high degree of writer and reader visibility" (p. 11), and she concludes (p. 12) that students "need to focus on the association between linguistic expressions and genre/style level/medium and on widening their repertoire of styles (e.g. formal/informal, impersonal/interactive)". This brings us to the next point.

## 9.6 Formality

It is a problem for students to maintain an appropriate style in the essays, for example with respect to sentence connection (**9.3.2**) and writer/reader visibility (**9.4-9.5**). Typically conversational features are illustrated in:

(114)   Another unpleasant thought is the thought of the victims. 'Victim' is not a pleasant word, ***is it?***

(115)   Bringing the money out of that vault is only part of the job ***you see***.

(116)   For some of us this may mean that it doesn't matter if we do something illegal, we won't get sentenced ***anyways***.

(117)   I do agree *in* [with] that statement, but ***anyway*** you have to use your imagination well in a lot of *datagames* [computer games] and *dataprogrammes* [computer programmes].

(118)   My last *practise* [practice] teacher was a lot younger than the *two first ones* [first two], and she had only been working as a teacher for four years. I found her more open-minded and not as restrictive as the other *practise* [practice] teachers I've had. She gave us quite a free hand to do what we wanted to do. This could also be because we were *3rd year student* [third-year students], with at least some experience. ***Anyway***, I relate it more to the fact that she was younger and not yet stuck in a certain structure, or way of acting as a teacher.

(119)   ***You see***, feminism to me has little or nothing to do with the right to vote, the right to choose, equal *pay-cheks* [pay-cheques], and the like. I take these things for granted, and at nineteen it's ***okay*** to be that way. At least that's what I like to think, ***I mean***. ***YES***, circumcision of women is clearly a very bad thing, as *is* [are] abusive husbands, obsessive boyfriends, date rape or just plain rape.

(120)   It is a *known fact* [well known] that the society we live in is constantly changing. Some might say for the better, ***whilst*** *some* [others] might say for the worse, ***anyhow*** it is still changing.

The tag question in (114) is a typically conversational feature, as is the sequence *you see* in (115). *Anyways* in (116) is a non-standard form which we do not expect to find in writing. The uses of *anyway* in the next two examples are informal and could be replaced here by *still* (117) and *in any case* (118). The next example contains a number of conversational features, including the capitalised *YES*, which is intended to express emphasis. Note, finally, the clash in (120) between the formal word *whilst* and the informal *anyhow*, here used in the same way as *anyway* in (118). There is also a comma splice, which is best avoided; see below.

## 9.7 Mechanics

By mechanics we understand the command of writing conventions such as spelling and punctuation. Spelling errors, including the use of the apostrophe, were taken up in Chapter 8 (**8.8**). A recurrent problem is the common use of **comma splices**, i.e. joining sentences by a comma instead of using a full stop or a conjunction. In (120) above it would be better to start a new sentence with *anyhow* or replace this word by the conjunction *but*. The following example contains a series of comma splices:

(121)   Making the army consist of only professional soldiers would be a mistake**,** it would leave the country unable to defend *it self* [itself]**,** *its* [it's] only 56 years since the world *where* [was] finished with World War 2**,** should we ignore that it ever happened? Believing that it would never happen again?

Here is a suggested rendering without comma splices and with errors corrected:

(121a) Making the army consist of only professional soldiers would be a mistake. It would leave the country unable to defend itself. It's only 56 years since the world was finished with World War 2. Should we ignore that it ever happened? ***Believing that it would never happen again?***

The suggested rendering retains a sentence fragment, without a subject plus verb, which could be rephrased in this way: *Should we believe it would never happen again?* Alternatively, we could join the fragment to the preceding sentence: *Should we ignore that it ever happened, believing that* [or: *in the belief that*] *it would never happen again?*[9] Here is another example of a sentence fragment:

(122)   *The prison* [Prison] keeps them away from *the society* [society]. ***Makes other people feel safe***. Anyway, you can *maximum be sentenced for 21 years* [be sentenced for a maximum of 21 years].

Suggested rendering: *It makes other people feel safe* or *... and makes other people feel safe*. Sentence fragments are commonly used in the NICLE material. Although they can be used to good effect, they are best avoided by student writers, or should at least used with caution.

More important still is the structuring of essays in terms of paragraphs. This goes beyond mere mechanics and has a lot to say for the comprehension of the text. NICLE writers have problems with paragraphing and often do not structure and mark paragraphs adequately. Although important, the question of paragraphing will not be taken up here. The same applies to an analysis of the essays in terms of argumentation. A proper treatment of these matters would require a different book.

## 9.8 Complex deficiency

Problems tend to go together as we have seen in many examples above. Often there is no simple way of correcting a sentence, and there is a need for a lot of rephrasing. These are some more examples, followed by suggested new renderings:

---

[9] Another possible rendering is: *Should we ignore that it ever happened? Believe that it would never happen again?*

(123)   We have this imagination that if a crime is done someone will be caught and punished. Either with fines or do jail time, and hopefully be rehabilitated not to do it again. But sadly that is not always the case, as we have seen to many times now.

(123a)  We believe that, if a crime has been committed, someone will be caught and punished, either with a fine or with a jail sentence, and we hope they will be rehabilitated so that they will not commit crimes again. But sadly that is not always the case, as we have seen many times now.

(124)   It would not likely be any society basing their rules of conduct solely on tacit agreement. Some form of formalized order must reign in order to assure that the individual's right is shown respect.

(124a)  There is probably no society which bases its rules of conduct solely on tacit agreement. There must be some formalized order to make sure that the individual's rights are respected.

(125)   Over the years, TV has become a global village, were interactions and belonging are central elements. An escape from reality, but still give us insight of it, are reasons enough why television is opium for the masses.

(125a)  Over the years TV has introduced us to a global village where interaction and contact are central elements. At the same time as it gives us insight, it allows us to escape from reality. This is a reason why it can be described as opium for the masses.

(126)   Idealistic struggles when loosing objectives to fight for, too often turns into exstremeties, and so also with the western feminists fights. Objects become more peculiar, and it now seems that the main object is to be superior to men.

(126a)  When one loses objectives to fight for, idealistic struggles too often turn into extremes, and this is also true of western feminists' fights. Objectives become more peculiar, and it now seems that the main objective is to be superior to men.

(127)   There are for sure important qualities to gain from university studies. Still, are they sufficient and satisfactory, considering what they could have been, if providing a more practical and contemporary approach, along with the present inclination? No doubt, students would win a lot through a closer cooperation with both co-students and lecturers.

(127a)  There are certainly important things to learn from university studies. But are they sufficient and satisfactory, considering what they could have been if universities offered a more practical and contemporary approach along with the present one? Students would no doubt gain a lot through closer cooperation with both fellow students and lecturers.

(128)   Speaking of technology as a general term, this is where the main problems lay. It is in this huge spectre of merchandise and inventions we find ourselves stuck with things that are not as important as the people that surround us. And by inventing this new stuff, to ease our own lives, we also create a need for something better. Therefore the circle of life spins once again: it seems, as if we cannot stop, we demand to have all of it and then some.

(128a)  Speaking of technology in general, this is where the main problems are. With the wide range of products and inventions, we find ourselves stuck with things that are not as

important as the people around us, and by these new means of easing our lives we also create a need for something better. Therefore the circle of life spins once more, and it seems as if we cannot stop. We want to have it all, and even more.

## 9.9 Analysing NICLE essays

The following scheme may be useful in analysing NICLE essays. Take up the most conspicuous points. Try to categorise your observations rather than just giving a list. In taking up different features it may be useful to say something about how pervasive they are. Are they pervasive or sporadic?

*Description*

Lexis:      form
            meaning
            collocation
            variation

Grammar:    morphology
            syntax
            complexity
            variation

Complex deficiency:  multiple problems, need to rephrase sentences

Text:       cohesion/coherence
            argumentation
            style:  formality
                    variation
                    reader-writer visibility

Mechanics:  paragraphing
            punctuation

*Explanation*

How can we account for the features observed? Influence from the mother tongue? General features of learner language?

*Evaluation*

Is the text successful? Does it make sense? Does it deal adequately with the topic? How could the text be improved?

## 9.10 Advice for learners

In her doctoral thesis, Linnarud (1986) compares compositions written by seventeen-year-old Swedish learners of English and essays written by native speakers. The study is similar to the comparison of ICLE and LOCNESS, but the material is much smaller and was not available in machine-readable form. Although the thesis focuses on lexis, Linnarud ends by listing some more general points of advice for student writers (p. 120f.). These are well worth heeding:

● Always think of composition as a means of expressing ideas and communicating a message. It is not primarily an exercise in using correct language, although correctness makes it easier, and more enjoyable, for the reader to understand the ideas you want to put across.

● Do not write too short a text. The reader must have enough material to be able to understand your message.

● Vary sentence construction and sentence length.

● Vary vocabulary by avoiding repetition of the same words and by using adjectives and adverbs to give more depth to your descriptions.

● Learn to use words in collocational frameworks rather than as single items. Very often there are specific adjectives that go with particular nouns. Learn them together as units. Learn and practise the use of multiword verbs.

● Finally, when you have finished your composition, go through it and check for errors such as concord and mixed tenses.

In Chapters 8 and 9 a lot of emphasis has been put on specific points where the learner may go wrong. These points are important, but we must not forget that the main purpose of writing is to express ideas and communicate a message. To quote a common saying, don't lose sight of the wood for the trees.

---

**9.11 Exercise**

Compare the original versions and the suggested renderings in **9.8**. What are the problems in the original versions? How can you account for the problems? Are the suggested new versions successful? Can you improve on them?

---

## *10. Problems and prospects*

### 10.1 The corpora

The use of corpora opens up new possibilities of comparing languages and of studying the characteristics of learner language, but there are also problems. The ENPC is fairly small and is of limited use for the study of words beyond the core vocabulary. For the analysis of grammatical features it normally requires a lot of manual intervention. Ideally, we would need a much larger corpus, preferably including annotation of grammatical features. We would also like to have a broader representation of text types, including comparable spoken material for the two languages.

The NICLE corpus is also insufficient for many types of studies. Above all, it only contains written argumentative essays by university students. We would like to have a far larger learner corpus, with a broader range of texts (including spoken material), ideally including material from different stages of learning. The material should preferably be annotated for error types and for grammatical and discourse features. Some developments along these lines have been planned or are in progress for other ICLE projects.

In spite of these limitations, the corpus material which we currently have at our disposal allows us to carry out many types of studies which may both be of theoretical interest and have practical applications. Some examples of applications will be illustrated below.

### 10.2 Lexicography

One of the most obvious applications is in bilingual lexicography. After the breakthrough in the 1980s, pioneered by the Cobuild project (Sinclair 1987), the use of monolingual corpora has brought about a revolution in monolingual lexicography. Although there have been some promising beginnings, a similar breakthrough is yet to come in the use of multilingual corpora.

It is a common experience that dictionaries fall short in the light of evidence from multilingual corpora. To take a couple of examples, even the biggest and most up-to-date English-Norwegian dictionaries do not give the user sufficient guidance for the translation of *spend* (**3.5.4**). Going in the other direction, readers are not told that *day* is the most common correspondence for *døgn* (**3.2**).

We cannot of course expect a traditional dictionary to contain all the context-sensitive information which we can draw from a multilingual corpus. However, given the possibilities of the electronic media, it is possible to connect dictionary and corpus. Ideally, we would also like to have a link with a grammar (cf. Johansson 1998: 21); see Figure 10.1. There are such links even in traditional dictionaries and grammars. Dictionaries contain grammatical information, such as word-class labels and indications of structural patterns. Grammars give lexical information, such as lists of uncountable nouns and verbs with particular complementation patterns. The importance of linking dictionary and grammar in a contrastive context is shown by the discussion of correspondences of *thing*, which has an important grammatical function (**3.4**).

Traditionally, the publications of lexicographers and grammarians were severely restricted by considerations of space, and exemplification was limited to instances out of context. We can now have an open-ended electronic dictionary which is linked to a corpus. Similarly, a grammar can be linked to a corpus. The dictionary and the grammar provide the description, and the corpus gives examples of language use in context. Ideally, we would like to have two-way links between grammar and corpus, dictionary and corpus, and grammar and dictionary.

This model is equally applicable to monolingual and cross-linguistic studies. Contrastive descriptions – lexical as well as grammatical – can be linked to corpora of parallel texts. To the user, dictionaries and grammars will no longer be collections of observations and examples out of context. They will be a guide to language use. Then we will be far away from the abstract system comparisons of early contrastive studies.
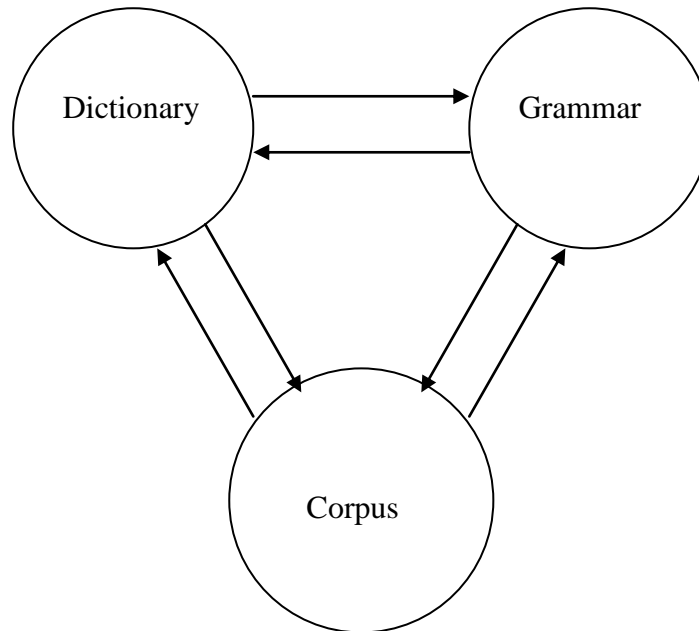


Figure 10.1 Dictionary, grammar and corpus: an integrated model

## 10.3 Translator training

It seems obvious that translators in training should be able to benefit greatly from studying existing translations as found in multilingual corpora:

> Existing translations contain more solutions to more translation problems than any other available resource. (Pierre Isabelle)[10]

> Corpora […] have more to offer translators than might at first sight seem apparent. Not only can they provide evidence for how words are used and what translations for a given word and phrase are possible, they also provide an insight into the process and nature of translation itself. (Hunston 2002: 128)

To take some examples, how do translators handle words which have no straightforward counterpart in the target language, such as Norwegian *døgn* (**3.2**) and English *mind* (**3.3**)? How are the problems posed by English complex sentence openings solved in translation into a V2 language (**5.2**)? How should we understand zero correspondence (**2.5**)? Is there compensation in the context, and if so how? Is any of the meaning lost? How should we interpret translation effects such as overuse and underuse (**2.8**)? The findings of corpus research provide ample material for discussion and consciousness-raising exercises.

---

[10] Quoted from: http://www.tsrali.com/help.cgi?topic=FAQ-why&UTLanguage=en&userName= (accessed in April, 2006)

It is a problem, however, that a translation corpus normally includes just one translation of each text, although we know that many other renderings would have been possible. To study the range of variation in translation we have compiled a smaller corpus where groups of translators have been given the same translation task.[11] This is how the first sentence in a short story by A. S. Byatt was handled by a group of ten professional translators:

(1)     IN THE MID-1980s, Bernard Lycett-Kean decided that Thatcher's Britain was uninhabitable, *a land of dog eat dog, lung-corroding ozone and floating money*, of which there was at once far too much and far too little. (Byatt, s. 1)

(1a)    Midt i åttiårene fant Bernard Lycett-Kean ut at Thatchers England ikke var til å leve i, *det var for mange bikkjer om beinet der, fullt av lungetærende ozon og flytende pund* som det både var for mye og for lite av. (transl. 1)

(1b)    På midten av nittenåttitallet fant Bernard Lycett-Kean ut at Thatchers Storbritannia var ulevelig, *et land i jungellovens tegn, med ozon som tærte på lungene og flytende penger*, som det var både altfor mye og altfor lite av. (transl. 2)

(1c)    Midtveis i 1980-åra slo Bernard Lycett-Kean fast at Thatchers England ikke var til å leve i, *et land av ulver, med lungetærende ozon og fri flyt av penger* som det på en gang var altfor mye og altfor lite av. (transl. 3)

(1d)    Mot midten av 1980-årene fant Bernard Lycett-Kean ut at Thatchers England ikke var til å leve i, *preget som det var av alles kamp mot alle, lungeetsende oson og flytende valuta* som det på samme tid var enten altfor mye eller altfor lite av. (transl. 4)

(1e)    Midt på åttitallet fant Bernard Lycett-Kean ut at Thatchers Storbritannia var ulevelig, *et nådeløst sted med lunge-etsende oson og raske penger*, som det samtidig var altfor mye og altfor lite av. (transl. 5)

(1f)    På midten av 1980-tallet var Bernard Lycett-Kean kommet til at Thatchers Storbritannia ikke lenger var til å bo i – *et land hvor livet var blitt det reneste bikkjeslagsmål, hvor ozonet åt opp lungene på folk, og pengene fløt i fritt fall*, altfor mye og altfor lite på en gang. (transl. 6)

(1g)    På midten av 80-tallet bestemte Bernard Lycett-Ken seg for at Thatchers England var ubeboelig. *Det var et land der alle var ute etter hverandre, med luft som etset opp lungene og fri flyt av penger*, som det forøvrig var både altfor mye og altfor lite av. (transl. 7)

(1h)    En gang på midten av 80-tallet fant Bernard Lycett-Kean ut at Thatchers Storbritannia ikke var et levelig sted. *Det var et land av alle mot alle, av lungetærende ozon og penger i fri flyt*, penger det både var altfor mye og altfor lite av. (transl. 8)

(1i)    Midt på 80-tallet bestemte Bernard Lycett-Kean seg for at Thatchers Storbritannia var et ubeboelig land, *hvor den sterkestes rett rådet, ozon etset lungene, og det var fri flyt av penger*, som det samtidig var altfor mye og altfor lite av. (transl. 9)

(1j)    Midt på 1980-tallet fant Bernard Lycett-Kean ut at Thatchers Storbritannia var blitt ubeboelig, *et barbarisk, kannibalistisk samfunn med dødelig oson og flytende penger*, som det både var altfor mye og altfor lite av på samme tid. (transl. 10)

Although all the translated versions differ greatly, the opening and the end of the sentences agree apart from minor differences in structure and word choice. The challenge is the noun phrase *a land of dog eat dog, lung-corroding ozone and floating money*. Two translators (7 and 8) started a new sentence, and most of them introduced one or more clauses. Such 'clause building' is commonly found in translation from English into Norwegian. The greatest problem is finding a way of handling the metaphorical *a land of dog eat dog*, which has no

---

[11] For information on this English-Norwegian Multiple-Translation Corpus, see Johansson (2007), Chapter 11.

exact counterpart in Norwegian. Some translators found a metaphorical expression (transl. 1, 2, 3, and 6), others opted for descriptive phrases or clauses (transl. 4, 7, 8, 9), and two reduced the phrase to adjectives which capture part of the meaning of *dog eat dog* (transl. 5 and 10). It is not easy to decide which is the most successful of the renderings – they have all been produced by experienced professional translators who had received prestigious prizes – but there is undoubtedly a lot of good material here for class discussion.

## 10.4 Foreign-language teaching

Corpora have many applications which are relevant to language teaching (see Figure 10.2). The findings of corpus studies can be used in the preparation of textbooks, grammars, dictionaries, and other teaching material. They can be used in syllabus design. They can be used in the training of teachers. They can be used in testing. These could all be called **indirect** applications.
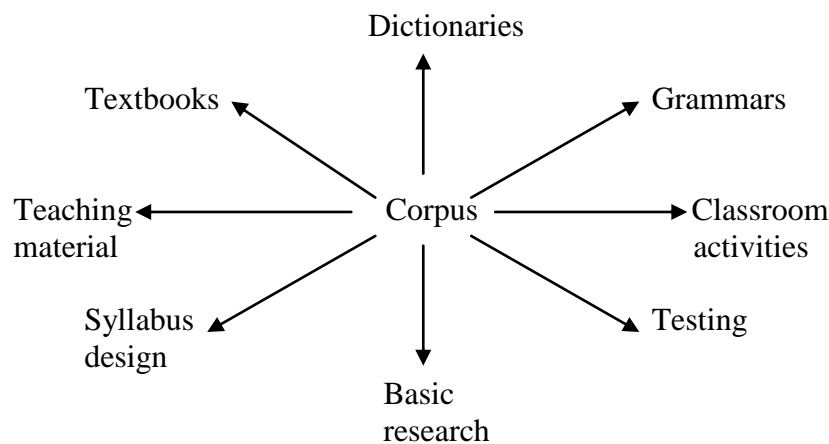


Figure 10.2 Uses of corpora of relevance for language teaching

  Corpora may also have **direct** applications in the classroom, at least at advanced levels. Learners can be trained to use corpora. Using a corpus like the ENPC they can be encouraged to study how words and phrases have actually been translated rather than just looking them up in a dictionary. Direct applications of a learner corpus like NICLE are harder to envisage, except perhaps as a means of training learners to spot problems in language use (and suggest possible improvements). Using corpora in the classroom requires that there are resources available (computers, corpora, programs). Above all, teachers are required who can guide learners in the use of corpora and in evaluating the material.
  The use of multilingual corpora and learner corpora may further be significant for basic research on language learning, as shown in Section **10.6**.

## 10.5 Combining learner corpora and multilingual corpora: An example

New possibilities are afforded by the combined use of learner corpora and multilingual corpora. These are some examples of the use of the phrase *the nature* taken from NICLE:[12]

(2)  From time to time a man should go to *the nature* to learn about himself and about his life.

---

[12] Note that these examples are given here exactly as they were produced, including errors.

(3)     […] to rest at a nice quiet place and enjoy the fresh air and the beauty of *the nature* after the work.

(4)     People want more and more to interfere in *the nature*.

(5)     *The naturre* still calls out but weaker and weaker. Once *the nature* stops calling it will have no strengh to recover.

Scandinavian students frequently write *the nature* with reference to nature in general. In fact, almost all occurrences of *the nature* in the Norwegian learner corpus are wrong. The explanation is mother-tongue influence. Like the other Scandinavian languages, Norwegian uses the definite form (signalled by a suffix) for generic reference where English requires a noun without a determiner: *naturen – nature*, *kunsten – art*, *historien – history*, *døden – death*, etc. (cf. Hasselgård *et al*. 2004: 115). After the problem has been pointed out, students can be directed to the ENPC to observe what translators have done. These are a few examples:

(6)     Utnyttelsen av det *naturen* hadde å gi, var og er grunnleggende. (AOH1)
        Exploiting what *nature* has to give was, and still is, a fundamental activity.

(7)     Knuser man druer, tar *naturen* hånd om resten. (KP1)
        If you crush the grapes, *nature* takes care of the rest.

(8)     Få har studert *naturen* mer intenst, med større vitebegjær og med skarpere øyne enn dyslektikeren Leonardo. (ANR1)
        Few persons have studied *nature* more intensely, with greater thirst for knowledge, and with a sharper eye than the dyslectic Leonardo.

(9)     He also came to understand *the nature* of courtiers in some degree. (RDA1)
        Han lærte også litt om hoffolks *natur*.

(10)    The painting was cut twelve times with a knife and from *the nature* of the damage it was deduced that the stabs and cuts were inflicted with great force. (JH1)
        Maleriet ble skåret tolv ganger med en kniv, og ut fra skadenes *natur* ble det sluttet at kuttene og flengene ble påført med stor kraft.

The first three examples show the standard translation pattern for Norwegian *naturen*. The last two illustrate how *the nature* is used in English. Observing the material in the corpus, the students will discover under what conditions this form is appropriate and how it is translated into Norwegian.

## 10.6 The integrated contrastive model

The importance of combining learner corpora and multilingual corpora extends beyond immediate teaching applications, as argued by Granger (1996) in her presentation of the integrated contrastive model; see Figure 10.3. The integrated contrastive model solves the most basic problem of applied contrastive analysis, viz. the problem that learning cannot be understood by a purely linguistic study. The question is not whether mother-tongue influence exists, but when and under what conditions. In contrastive analysis (CA) we compare original texts (OL vs. OL) across languages or source- and target-language texts (SL vs. TL). From the comparison we predict potential difficulties for learners. By contrastive interlanguage analysis (CIA), comparing native and learner language use (NL vs. IL) and different groups of learners (IL vs. IL), we find out what the actual problems are and to what extent they are shared irrespective of the mother tongue of the learners. Here the approach is diagnostic. Learner-

language features may be traced back to the mother tongue, but they may also reflect general strategies of learning or have other causes, as indicated by the diverging arrows in Figure 10.3.
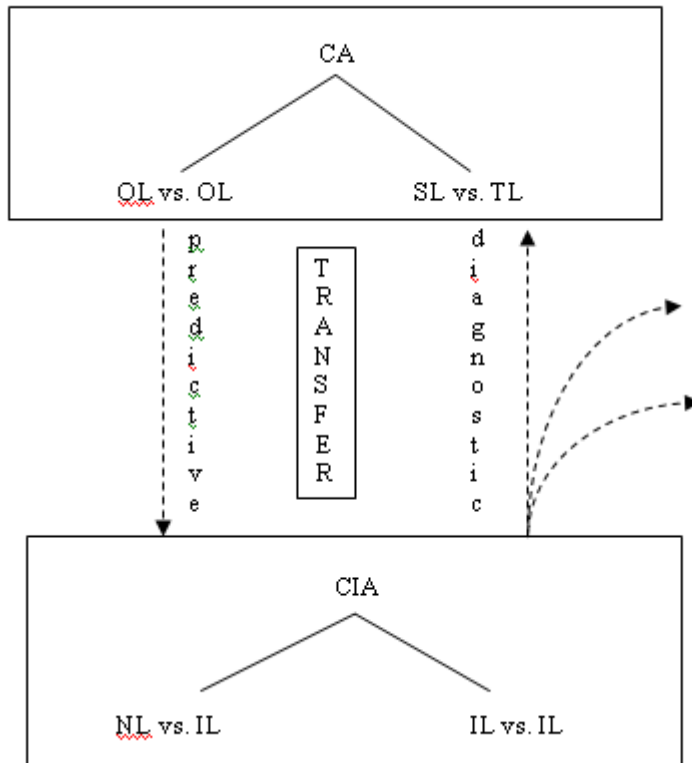
Figure 10.3. The integrated contrastive model (quoted from Gilquin 2000/2001: 100, based on Granger 1996: 47)

Combining learner corpora and multilingual corpora is a new development. Though there are some promising beginnings (e.g. Gilquin 2000/2001), there is far more to do, with respect to both corpus building and research. This is well worth the effort, because the integrated contrastive model holds a key to the understanding of foreign-language acquisition. To what extent is it guided by the mother tongue? To what extent does it reflect general processes of language learning?

## 10.7 Where do we go?

As I hope to have shown, there is a lot we can do with the corpora which are currently available. We learn about languages and their relationships. We learn about translation. We get new insight into learner language. But the use of multilingual and learner corpora is still fairly new, and there is far more we need to do (**10.1**). We need more corpora. We need better corpora. We need to learn to exploit corpora in an optimal manner.

Corpora must be used with caution. Corpus findings must be evaluated against possible shortcomings of the available material. The corpus user must always be prepared to move beyond the corpus, extending the corpus or using other means of exploration (intuition, elicitation, informal observation), because using corpora is just one means of gaining insight into language structure and use.

# References

Ädel, A. 2006. *Metadiscourse in L1 and L2 English*. Amsterdam & Philadelphia: Benjamins.

Aijmer, K. 2002. Modal adverbs of certainty and uncertainty in an English-Swedish perspective. In Hasselgård *et al*. (2002), 97-112.

Aijmer, K. and B. Altenberg. 1996. Introduction. In Aijmer *et al*. (1996), 11-16.

Aijmer, K., B. Altenberg, and M. Johansson (eds). 1996. *Languages in Contrast. Papers from a symposium on text-based cross-linguistic studies, Lund 4-5 March 1994*. Lund Studies in English 88. Lund: Lund University Press.

Aijmer, K. and H. Hasselgård (eds). 2004. *Translation and Corpora. Selected papers from the Göteborg-Oslo symposium 18-19 October 2003*. Gothenburg Studies in English 89. Göteborg: Acta Universitatis Gothoburgensis.

Aijmer, K. and A.-M. Simon-Vandenbergen. 2003. The discourse particle *well* and its equivalents in Swedish and Dutch. *Linguistics* 41: 1123-1161.

Altenberg, B. 1998. Connectors and sentence openings in English and Swedish. In Johansson and Oksefjell (1998), 115-143.

Altenberg, B. 1999. Adverbial connectors in English and Swedish: Semantic and lexical correspondences. In Hasselgård and Oksefjell (1999), 249-268.

Altenberg, B. and K. Aijmer. 2000. The English-Swedish Parallel Corpus: A resource for contrastive research and translation studies. In C. Mair and M. Hundt (eds), *Corpus Linguistics and Linguistic Theory. Papers from the Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20), Freiburg im Breisgau, 1999*. Amsterdam & Atlanta: Rodopi, 15-33.

Altenberg, B. and S. Granger (eds). 2002. *Lexis in Contrast: Corpus-based approaches*. Amsterdam & Philadelphia: Benjamins.

Altenberg, B. and M. Tapper. 1998. The use of adverbial connectors in advanced Swedish learners' written English. In Granger ed. (1998), 80-93.

Baker, M. 1993. Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, and E. Tognini-Bonelli (eds), *Text and Technology. In honour of John Sinclair*. Amsterdam & Philadelphia: Benjamins, 233-250.

Baker, M. 1995. Corpora in translation studies: An overview and some suggestions for future research. *Target* 7/2: 233-250.

Bjerga, T. D. 1998. Continuative and habitual aspect in English and Norwegian, with special reference to the English verb *keep* and the Norwegian verb *pleie*. *Hovedfag* thesis. Department of British and American Studies, University of Oslo.

Boström Aronsson, M. 2003. On clefts and information structure in Swedish EFL writing. In S. Granger and S. Petch-Tyson (eds), *Extending the Scope of Corpus-based Research: New applications, new challenges*. Amsterdam: Rodopi, 197-210.

Carl, M. and A. Way (eds). 2003. *Recent Advances in Example-based Machine Translation*. Dordrecht: Kluwer.

Chesterman, A. 1998. *Contrastive Functional Analysis*. Amsterdam & Philadelphia: Benjamins.

Chocholousová, B. 2003. English *thing* vs. Norwegian *ting*: How well do they correspond? Term paper, Contrastive analysis and learner language, Spring 2003. Department of British and American Studies, University of Oslo.

Davidsen-Nielsen, N. 1996. Discourse particles in Danish. In E. Engberg-Pedersen, M. Fortescue, P. Harder. L. Heltoft, and L. Falster Jacobsen (eds), *Content, Expression and Structure. Studies in Danish functional grammar*. Amsterdam & Philadelphia: Benjamins, 283-314.

de Vries, G. T. 2001. English tag questions and their equivalents in Norwegian and Dutch. *Hovedfag* thesis. Department of British and American Studies, University of Oslo.

Dyvik, H. 1998. A translational basis for semantics. In Johansson and Oksefjell (1998), 51-86.

Dyvik, H. 2005. Translations as a semantic knowledge source. In *Proceedings of the Second Baltic Conference on Human Language Technologies*. Tallinn: Institute of Cybernetics at Tallinn University of Technology, Institute of the Estonian Language, 27-38.

Ebeling, J. 1998. The Translation Corpus Explorer: A browser for parallel texts. In Johansson and Oksefjell (1998), 101-112.

Ebeling, J. 2000. *Presentative Constructions in English and Norwegian: A corpus-based contrastive study*. Acta Humaniora 68, Faculty of Arts, University of Oslo. Oslo: Unipub forlag.

Eia, A.-B. 2006. The use of linking adverbials in Norwegian advanced learners' written English. Master's thesis. Department of Literature, Area Studies and European Languages, University of Oslo.

Elsness, J. 2000/2001. A contrastive look at the present/preterite opposition in English and Norwegian. *Languages in Contrast* 3/1: 3-40.

Engen, J. H. 2001. *Og isn't always and*: From coordination to subordination in English translations of Norwegian texts. *Hovedfag* thesis. Department of British and American Studies, University of Oslo.

Faarlund, J. T., S. Lie, and K. I. Vannebo. 1997. *Norsk referensegrammatikk*. Oslo: Universitetsforlaget.

Fabricius-Hansen, C. 1998. Informational density and translation, with special reference to German – Norwegian – English. In Johansson and Oksefjell (1998), 197-234.

Færch, C., K. Haastrup, and R. Phillipson (in collaboration with E. Glahn and K. Bæk Kristiansen). 1984. *Learner Language and Language Learning*. Copenhagen: Gyldendal.

Firbas, J. 1992. *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: Cambridge University Press.

Fretheim, T. 2005. Politeness in Norway: How can you be polite and sincere? In L. Hickey and M. Stewart (eds), *Politeness in Europe*. Clevedon: Multilingual Matters, 145-158.

Fries, Ch. C. 1945. *Teaching and Learning English as a Foreign Language*. Ann Arbor: University of Michigan Press.

Gellerstam, M. 1996. Translations as a source for cross-linguistic studies. In Aijmer *et al*. (1996), 53-62.

Gilquin, G. 2000/2001. The Integrated Contrastive model: Spicing up your data. *Languages in Contrast* 3/1: 95-123.

Granger, S. 1996. From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In Aijmer *et al*. (1996), 37-51.

Granger, S. 1998. The computer learner corpus: A versatile new source of data for SLA research. In Granger ed. (1998), 3-18.

Granger, S. (ed.). 1998. *Learner English on Computer*. London: Longman.

Gundel, J. 2002. Information structure and the use of cleft sentences in English and Norwegian. In Hasselgård *et al*. (2002), 113-128.

Gundersen, K. 2004. Norwegian preposition + *at*-clause and its correspondences in English. In Aijmer and Hasselgård (2004), 113-127.

Guslund, I. M. 1996. The Norwegian modal verb *skal* and its equivalents in English: A contrastive analysis. *Hovedfag* thesis. University of Oslo.

Halliday, M. A. K. 2004. *An Introduction to Functional Grammar*. 3rd ed. Revised by Chr. M. I. M. Matthiessen. London: Arnold.

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

Hasselgård, H. 1997. Sentence openings in English and Norwegian. In M. Ljung (ed.), *Corpus-based Studies in English. Papers from the Seventeenth International Conference on English Language Research on Computerized Corpora*. Amsterdam: Rodopi, 3-20

Hasselgård, H. 1998. Thematic structure in translation between English and Norwegian. In Johansson and Oksefjell (1998), 145-168.

Hasselgård, H. 2000. English multiple themes in translation. In A. Klinge (ed.), *Contrastive Studies in Syntax*. Copenhagen Studies in Language 25. Frederiksberg: Samfundslitteratur, 11-38.

Hasselgård, H. 2004a. Thematic choice in English and Norwegian. *Functions of Language* 11/2: 187-212.

Hasselgård, H. 2004b. Spatial linking in English and Norwegian. In Aijmer and Hasselgård (2004), 163-188.

Hasselgård, H. 2005. Theme in Norwegian. In K. L. Berge and E. Maagerø (eds), *Semiotics from the North. Nordic approaches to systemic functional linguistics*. Oslo: Novus, 35-47.

Hasselgård, H. Forthcoming. Thematic choice and expressions of stance in English argumentative texts by Norwegian learners. In K. Aijmer (ed.), *Corpora and Language Teaching*. Amsterdam & Philadelphia: Benjamins.

Hasselgård, H. and S. Oksefjell (eds). 1999. *Out of Corpora. Studies in honour of Stig Johansson*. Amsterdam & Atlanta, GA: Rodopi.

Hasselgård, H., S. Johansson, B. Behrens, and C. Fabricius-Hansen (eds). 2002. *Information Structure in a Cross-linguistic Perspective*. Amsterdam & Atlanta, GA: Rodopi.

Hasselgård, H., S. Johansson, and P. Lysvåg. 2004. *English Grammar: Theory and use*. 3rd ed. Oslo: Universitetsforlaget.

Hasselgren, Angela. 1994. Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary, *International Journal of Applied Linguistics* 4: 237-259.

Hoey, M. and D. Houghton. 1998. Contrastive analysis and translation. In M. Baker (ed., assisted by K. Malmkjær), *Routledge Encyclopedia of Translation Studies*. London & New York, 45-49.

Hofland, K. 1996. A program for aligning English and Norwegian sentences. In S. Hockey, N. Ide, and G. Perissinotto (eds), *Research in Humanities Computing*. Oxford: Oxford University Press, 165-178.

Hofland, K. and S. Johansson. 1998. The Translation Corpus Aligner: A program for automatic alignment off parallel texts. In Johansson and Oksefjell (1998), 87-100.

House, J. and G. Kasper. 1981. Politeness markers in English and German. In F. Coulmas (ed.), *Conversational Routine. Explorations in Standardized Communication Situations and Prepatterned Speech*. The Hague: Mouton, 157-185.

Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hunston, S. and G. Francis. 2000. *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam & Philadelphia: Benjamins.

Jakobson, R. 1959. On linguistic aspects of translation. In R. A. Brower (ed.), *On Translation*. Cambridge, Mass: The M.I.T. Press, 232-239.

James, C. 1980. *Contrastive Analysis*. London: Longman.

James, C. 1998. *Errors in Language Learning and Use: Exploring error analysis*. London: Longman.

Johannessen, T. A. 2003. Non-restrictive relative clauses in the English-Norwegian Parallel Corpus. Term paper, Contrastive analysis and learner language, Spring 2003. Department of British and American Studies, University of Oslo.

Johansson, M. 2002. *Clefts in English and Swedish: A contrastive study of* it-*clefts and* WH-*clefts in original texts and translations*. Doctoral dissertation. Department of English, Lund University.

Johansson, S. 1998. On the role of corpora in cross-linguistic research. In Johansson and Oksefjell (1998), 3-24.

Johansson, S. 2004. What is a person in English and Norwegian? In Aijmer and Hasselgård (2004), 69-85.

Johansson, S. 2007. *Seeing through Multilingual Corpora: On the use of corpora in contrastive studies*. Amsterdam & Philadelphia: Benjamins.

Johansson, S., J. Ebeling, and S. Oksefjell. 1996. Coding and aligning the English-Norwegian Parallel Corpus. In Aijmer *et al.* (1996), 87-112.

Johansson, S. and K. Hofland. 1994. Towards an English-Norwegian parallel corpus. In U. Fries, G. Tottie, and P. Schneider (eds), *Creating and Using English Language Corpora*. Amsterdam & Atlanta, GA: Rodopi, 25-37.

Johansson, S. and B. Løken. 1997. Some Norwegian discourse particles and their English correspondences. In C. Bache and A. Klinge (eds), *Sounds, Structures and Senses*.

*Essays presented to Niels Davidsen-Nielsen on the occasion of his sixtieth birthday*. Odense: Odense University Press, 149-170.

Johansson, S. and H. Stavestrand. 1987. Problems in learning – and teaching – the progressive form. In I. Lindblad and M. Ljung (eds), *Proceedings from the Third Nordic Conference for English Studies*. Stockholm Studies in English 73. Vol. 1. Stockholm: Almqvist & Wiksell, 139-148.

Kennedy, G. 1992. Preferred ways of putting things with implications for language

Teaching. In J. Svartvik (ed.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*. Berlin & New York: Mouton de Gruyter, 335-373.

Kenny, D. 1998. Equivalence. In M. Baker (ed., assisted by K. Malmkjær), *Routledge Encyclopedia of Translation studies*. London & New York: Routledge, 77-80.

Krzeszowski, T. P. 1995. *Early Contrastive Studies in England (15$^{th}$-18$^{th}$ centuries)*. Gdansk: Wydawnictwo Uniwersytetu Gdanskiego.

Lado, R. 1957. *Linguistics Across Cultures: Applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.

Lakoff, G. and M. Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Lauridsen, K. 1996. Text corpora in contrastive linguistics: Which type of corpus for which type of analysis? In Aijmer *et al*. (1996), 63-71.

Laviosa, S. 1998. The corpus-based approach: A new paradigm in translation studies. *META* 43: 474-479.

Linnarud. M. 1986. *Lexis in Composition: A performance analysis of Swedish learners' written English*. Lund Studies in English 74. Lund: CWK Gleerup.

Løken, B. 2007. *Beyond Modals: A corpus-based study of English and Norwegian expressions of possibility*. Acta Humaniora 296, Faculty of Arts, University of Oslo. Oslo: Unipub forlag.

Mathesius, V. 1975. *A Functional Analysis of Present-day English on a General Linguistic Basis*. Transl. L. Dusková, ed. J. Vachek. Prague: Academia.

Petch-Tyson, S. 1998. Writer/reader visibility in EFL written discourse. In Granger ed. (1998), 107-118.

Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Ringbom, H. 1987. *The Role of the First Language in Foreign Language Learning*. Clevedon: Multilingual Matters.

Ringbom, H. 1994. Contrastive analysis. In R. E. Asher and J. M. Y. Simpson (eds), *Encyclopedia of Linguistics*, Vol. 2. Oxford: Pergamon Press, 737-742.

Ringbom, H. 1998. Vocabulary frequencies in advanced learner English: A cross-linguistic approach. In Granger ed. (1998), 41-52.

Schmid, H.-J. 2000. *English Abstract Nouns as Conceptual Shells. From corpus to cognition*. Berlin & New York: Mouton de Gruyter.

Sinclair, J. (ed.).1987. *Looking Up: An account of the COBUILD Project in Lexical Computing*. London: Harper Collins.

Sønsterudbråten, T. 2003. Words for night and day in English and Norwegian. Term paper, Contrastive analysis and learner language, Spring 2003. Department of British and American Studies, University of Oslo.

Spalatin, L. 1969. Approach to contrastive analysis. In *The Yugoslav Serbo-Croatian - English Contrastive Project B. Studies 1*. Institute of Linguistics, University of Zagreb, 26-36.

Sperberg-McQueen, M. C. M. and L. Burnard (eds). 1994. *Guidelines for Electronic Text Encoding and Interchange. TEI P3*. Chicago & Oxford: Association for Computers and the Humanities / Association for Computational Linguistics / Association for Literary and Linguistic Computing.

Teubert, W. 1996. Comparable or parallel corpora? *International Journal of Lexicography* 9: 238-264.

Thagg, U. 1985. *The Sweet Sound of Concord: A study of Swedish learners'concord problems in English*. Lund Studies in English 73. Lund: CWK Gleerup.

Thomas, Jenny. 1983. Cross-cultural pragmatic failure, *Applied Linguistics* 4: 91-112.

Thompson, Geoff. 2004. *Introducing Functional Grammar*. 2nd ed. London: Arnold.

Thormodsæter, Ø. 2006. Aspects of the progressive: A comparative analysis of the English progressive aspect and its correspondences in Norwegian. Master's thesis. Department of Literature, Area Studies and European Languages, University of Oslo.

Tonne, I. 2001. *Progressives in Norwegian and the Theory of Aspectuality*. Acta Humaniora 104, Faculty of Arts, University of Oslo. Oslo: Unipub forlag.

Toury, G. 1980. *In Search of a Theory of Translation*. Tel Aviv: Porter Institute.

Toury, G. 1995. *Descriptive Translation Studies and Beyond*. Amsterdam & Philadelphia: Benjamins.

van Roey, J. 1990. *French-English Contrastive Lexicology: An introduction*. Louvain-la-Neuve: Peeters.

Véronis, J. (ed.). 2000. *Parallel Text Processing. Alignment and use of translation corpora*. Dordrecht etc.: Kluwer.

Vestre, A. B. 2003. Supplementive *ing*-clauses and their correspondences in Norwegian. Term paper, Contrastive analysis and learner language, Spring 2003. Department of British and American Studies, University of Oslo.

Viberg, Å. 1996. Cross-linguistic lexicology. The case of English *go* and Swedish *gå*. In Aijmer *et al*. (1996), 151-182.

Virtanen, T. 1998. Direct questions in EFL written discourse. In Granger ed. (1998), 94-106.