

Universität des Saarlandes - Philosophische Fakultät II  
FR 4.6 Angewandte Sprachwissenschaft  
sowie Übersetzen und Dolmetschen

## **Abschlussarbeit**

zur Erlangung des akademischen Grades einer  
Diplom-Dolmetscherin

# **Identifying English Gerunds and their Translation Equivalents in an English- German Translation Corpus**

Vorgelegt von:  
Katrin Menzel

Erstgutachter: Dr. Stella Neumann, PD  
Zweitgutachter: Prof. Dr. Erich Steiner  
Saarbrücken, 10. Oktober 2010

## **Abstract**

The gerund is a peculiarity of English grammar that has no exact equivalent in German. It displays both nominal and verbal properties. Finding a clear-cut definition of gerunds is challenging and still the subject of debate. This diploma thesis aims to identify English gerunds, to compare them to other -ing forms and to take into account in what historical context gerunds emerged. It briefly compares them to gerund-like forms in other Indo-European languages and looks at typical patterns and functions that are associated with English gerunds. By providing a systematic analysis of corpus passages containing gerunds that are identified via corpus queries and by analysing their frequencies and their German equivalents in an English-German translation corpus, this thesis aims to identify the most common strategies that translators use in different textual registers for translating gerunds into German.

## CONTENTS

1	Introduction.....	5
2	A contrastive and diachronic perspective on gerunds .....	9
2.1	Gerund-like forms in other languages.....	9
2.2	A diachronic account of the English gerund.....	13
2.2.1	The transition from Old English to Middle English.....	13
2.2.2	The development of the English gerund.....	15
3	A typology of <i>-ing</i> forms .....	20
3.1	The <i>-ing</i> suffix in common nouns .....	20
3.2	The <i>-ing</i> suffix in present participles .....	22
3.3	The <i>-ing</i> suffix in gerunds .....	26
4	A more detailed description of English gerunds .....	29
4.1	Gerunds as a mixed category .....	29
4.2	The structure of English gerunds (Poss-ing, Acc-ing, PRO-ing, Ing-of).....	34
4.3	Analysis of Acc-ing and Poss-ing gerunds as compared to Ing-of constructions.....	37
5	German translation equivalents for English gerunds.....	43
6	The analysis of gerunds in electronic corpus data.....	50
6.1	Corpus linguistics and corpus-based translation studies.....	50
6.2	The CroCo corpus .....	54
6.3	Some previous corpus-based studies on gerunds.....	59
6.4	Gerunds and their translations the parallel corpus <i>CroCo</i> ..	70
6.4.1	Developing a query mechanism.....	70
6.4.2	Investigating certain gerund patterns.....	78
6.4.2.1	Ing-of gerunds and prepositions + gerunds .....	78
6.4.2.2	Manual disambiguation and technical problems .....	83
6.4.2.3	Query results and analysis .....	86
6.4.2.4	Particular characteristics of individual registers.....	104
7	Conclusions and suggestions for further research .....	106
	References.....	108

## List of Figures

Fig. 1: Major periods of the English language .....	13
Fig. 2: <i>-ing</i> suffixes according to the OED.....	20
Fig. 3: Expressions followed by gerunds.....	27
Fig. 4: The scale of nouniness according to Ross.....	31
Fig. 5: The CroCo corpus .....	54
Fig. 6: Profiles of the English original registers .....	57
Fig. 7: Profiles of the German original registers .....	58
Fig. 8: Possessive determiners and objective pronouns + V- <i>ing</i> ..	61
Fig. 9: Frequencies of German translation options for gerunds in corpus data .....	64
Fig. 10: POS-distribution of <i>-ing</i> forms in E2G_ESSAY_EN.....	72
Fig. 11: Query excerpt from E2G_FICTION_EN.....	73
Fig. 12: Query results Ing-of gerunds in CroCo.....	81
Fig. 13: Query results preposition + gerund in CroCo .....	82
Fig. 14: German equivalents of Ing-of gerunds in CroCo (E2G)..	87
Fig. 15: German equivalents of Ing-of gerunds in CroCo (G2E)..	88
Fig. 16: German equivalents of prep. + gerunds in CroCo (E2G)	89
Fig. 17: German equivalents of prep. + gerunds in CroCo (G2E)	90
Fig. 18: Highest / lowest gerund frequencies in CroCo registers.	91
Fig. 19: Tokens and sentences per register.....	101
Fig. 20: Ratio of gerunds per tokens and sentences .....	102

# <sup>1</sup>1 Introduction

This thesis investigates a particular puzzling grammatical phenomenon of the English language – the gerund. The grammatical analysis of gerunds raises interesting contradictions and inconsistencies that cannot be rationalised away easily and the problem is worthy of debate. Gerunds in particular highlight the compromise required between grammatical analysis and purity of description with every day use and language development. The gerund is a peculiarity of English grammar that has no exact equivalent in German. It displays both nominal and verbal properties. Finding a clear-cut definition of gerunds is challenging and still the subject of debate.

This study discusses the use of gerunds and difficulties associated with their classification. Linguists and grammarians are interested in the grammatical classification of gerunds, in their historical development and in their functions and syntactic features. Translators, teachers who design courses and language learners must be able to understand the functions of gerunds and language contrasts between English and their native tongue. Finally, understanding the syntax and the use of gerunds is important for native speakers of English who want to communicate efficiently.

---

<sup>1</sup> Note: The title of the thesis has been changed before publication. The original title of the submitted thesis was *Identifying and Translating English Gerunds, with Particular Reference to an English-German Translation Corpus*.

First, this thesis briefly takes account of gerund-like forms in other Indo-European languages and summarizes syntactic diachronic changes between Old English and Middle English when gerunds established themselves as an indispensable part of the English grammatical system. Then it provides a classification of English *-ing* forms. English has some distinct morphemes sharing the phonological shape *-ing*. The *-ing* suffix has a wide range of functions and uses. The distinct morphological property of this suffix is that it attaches to verbs and the resulting lexical item may be of different part-of-speech categories – it is a category-neutral affix. Additionally, the English *-ing* affix is ambiguous between derivational and inflectional use. It can, for example, transform a verb into a noun or appear in present participles and gerunds.

Nonetheless, the distinction between the different categories of *-ing* forms and particularly the definition of gerunds are still open questions, highly debated by grammarians. These forms apparently are grammatical phenomena combining inherent contradictory features. The gerund merits discussion as it highlights several challenges faced by grammarians. Grammarians have been thinking in terms of lexical categories (e.g. nouns and verbs) since the very beginning. Some constructions however appear to combine the properties of two different categories simultaneously. The gerund displays both nominal and verbal properties and belongs to a mixed category in grammar due to the historical development of English.

The discussion of synchronic and language-contrastive aspects with regard to gerunds as well as the discussion of gerund frequencies in different text types and registers contributes to understanding and developing translation strategies when translating English passages containing gerunds into other languages that do not have a grammatical pattern as an exact equivalent of the structure from the source text. This study focuses on the English-German language pair and gives an overview of approximately equivalent German structures for English gerunds and explains under what condition one variant is preferred over the others in German. The corpus study that is described in this second part of this thesis is different from several previous corpus-based studies on selected aspects of gerunds by not following their primary focus on historical aspects and the emergence of gerunds in early periods of English in diachronic monolingual corpus data.

Among the few studies which have investigated the translation of gerunds in English-German parallel corpus data, Mehl's studies (1995, 1996) are worth noting. His work focuses on the development of translation algorithms for gerunds in Machine Translation systems. For his purposes, he chose a translation corpus that is rather small and consists only of one text – the text of an English grammar book and its German translation. His studies raise some interesting points by quantitatively analysing the options for translating gerunds into German. This thesis will expand on Mehl's results and observations by testing them against a larger and more representative corpus comprised of different

texts and registers and by trying to develop corpus queries that lead to an output that need less manual disambiguation. The analysis of gerunds and their German equivalents in an English-German translation corpus in this thesis contributes to the discussion of typical strategies for translating English clauses containing gerunds into German.



## 2 A contrastive and diachronic perspective on gerunds

### 2.1 Gerund-like forms in other languages

Gerund-like forms have existed and still exist in various languages. For instance, we already find a past gerund form in Sanskrit.

(1) अर्जुन गजं दृष्ट्वा वनं गच्छति  
arjuno gajaṃ dr̥ṣṭvā vanaṃ gacchati <sup>2</sup>

After seeing / Having seen the elephant, Arjuna goes to the forest.

One possible explanation for the emergence of gerund-like forms in Indo-European languages is that they have arisen from the fossilised instrumental case of verbal nouns (Tikkanen, 1987). Another view is that they developed from participles, particularly in Romance, Slavic and Baltic languages. These gerund-like forms acquired the verbal functions of the active participle whereas the latter became more like an adjective. In Greek, for example, it seems as if the uninflected Modern Greek active gerund, which is actually quite different from the English gerund, has developed

---

<sup>2</sup> <http://www.ibiblio.org/sanskrit/fundamentals/voices/gerund> (last accessed 8 August 2010)

from the inflected Ancient Greek active participle (Manolessou, 2005).

The so-called gerund or *gerundium* in Latin expresses a generalised or uncompleted action. It is a defective verbal noun with the following forms:

genitive singular (sg.)	-ndi
accusative sg.	-ndum
dative / ablative sg.	-ndo

(genitive)

- (2) studium linguam discendi – the desire of learning a language
- (3) servandi gratia – for the sake of rescuing

(dative)

- (4) aqua utilis bibendo – water good for drinking
- (5) legendo dare operam – to try hard with learning

(accusative)

- (6) idoneus ad legendum – suitable for reading

(ablative)

- (7) docendo discere – learning by teaching

These types of Latin gerunds have no plural form and there is no nominative form either; instead the infinitive is used (Jasanoff, 2006). In Latin, we also find the similar-looking *gerundive* – the future passive participle – which functions as a verbal adjective. It expresses the necessity of the action to be performed, and has the suffix *-ndus, -a, -um* (cf. [8]).

- (8) Carthago *delenda* est.  
Carthage is to be / must be destroyed.

French has a form of the present participle after ‘*en*’ (9) which is called the *gérondif*. In many cases, this form rather resembles the English participle in abridged adverbial clauses than the English gerund (cf. 10).

- (9) Ils pensaient qu'*en travaillant* plus ils gagneraient plus.  
They thought they would earn more by working more.
- (10) *En attendant* l'avion, il a ouvert sa valise.  
Waiting for the plane, he opened his suitcase.

Many French prepositions are not followed by a present participle / *gérondif*, but by an infinitive, cf. (11) and (12).

- (11) sans savoir qu'il était russe – without knowing that he was Russian
- (12) avant de partir – before leaving

The Portuguese gerúndio (similar to the Spanish and Italian gerundio) is used like the progressive or a present participle (cf. [13] and [14]). It has almost entirely lost its verbal nature and has become an adjective or substantive in examples such as *doente* – ‘ill’ / ‘patient’.

- |      |                                                     |
|------|-----------------------------------------------------|
| (13) | Ela ainda está dormindo.<br>She is still sleeping.  |
| (14) | Estou escrevendo uma carta.<br>I am writing a card. |

In sum, gerund-like forms appear in various languages as non-finite verb forms. However, the English gerund is a unique form with no exact equivalent in any other language. The functions of gerund-like forms in other languages often correspond to the functions of English participles and not to those of English gerunds.

## 2.2 A diachronic account of the English gerund

### 2.2.1 The transition from Old English to Middle English

The English gerund emerged during the transition from Old English to Middle English, a period of tremendous grammatical changes for the Anglo-Saxon language in general.

Old English (OE, 450-1100)
Middle English (ME, 1100-1500)
Early Modern English (EModE 1500-1800)
Late Modern English (LModE, 1800-Present)

Fig. 1: Major periods of the English language (cf. Algeo, J., 2010: 10)

From the 8th to the 11th century, the northern British Isles were influenced by Scandinavian invasions. In the far north, Gaelic prevailed and Old Scots survived, but England was partitioned and linguistically separated although commingling. As part of the Treaty of Wedmore, a boundary was drawn across England from London to the river Mersey. South of this line, the laws and customs would be those of the English, under the rule of the King of Wessex. The land to the north and east of this line would be under Viking rule, with Scandinavian laws and customs. This Viking part of England became known as the Danelaw. Viking settlements were founded and coexisted with the Germanic

kingdoms in Britain. OE and Scandinavian varieties had many words in common and might have been mutually intelligible. It could have been the case that the Scandinavian influence triggered the morphological and syntactic changes from OE to ME. After the Norman Conquest (1066), Anglo-Norman (and later Anglo-French, which was closer to Parisian French) became the main administrative language and was spoken by aristocratic families while Latin remained the main written language. Only gradually did Middle English gain more influence after 1300 when it re-emerged as a language used in literature, court and church.

English underwent many changes during that time. It changed, for instance, from a heavily synthetic language to a more analytic language. More and more inflectional features disappeared and were substituted by new syntactic or periphrastic constructions. Old English differentiated between four cases and still had the category of grammatical gender. The language has lost grammatical gender, adjectival and article inflection and its original rich verb inflection system. Auxiliaries, *will*-future and periphrastic *do* were developed and nominal case inflection has largely been substituted by prepositional constructions. The continuing loss of case inflection was accompanied with an increasingly fixed word order (SVO).

## 2.2.2 The development of the English gerund

In Old English, *-ing* occurred as part of word stems in nouns:

(15) pening / penig (penny)

(16) hring (ring)

*-ing* and *-ung* also combined with word-stems as a suffix.

(17) cyning (king) from OE cynn (family, race)

(18) wicing (Viking) from OE wic (village, camp)

(19) bletsunga, bledsunge (blessing) from OE bletsian, bledsian (consecrate, give thanks)<sup>3</sup>

In *Beowulf*, the most famous epic from OE literature, we already find some verbal nouns, e.g. *trimming* (strengthening) from *trymman* (strengthen) and *gemeting* (meeting) from *metan* (meet) (Cho, 1985).

The first gerunds in English developed probably from verbal nouns ending in *-ung* / *-ing*. They occurred for the first time towards the end of the Old English period with strong feminine declension (Curme, 1980). In that sense, they were similar to the form *-ung* in Modern German. The form *-yng* also occurred during the transition period from *-ung* to *-ing* in ME Northern, Midland

---

<sup>3</sup> Online Etymology Dictionary <http://www.etymonline.com/> (last accessed 8 August 2010)

and Southern dialects. The transition from *-ung* to *-ing* probably happened because *-ung* was very similar to *-ing* in its pronunciation both in Anglo-Saxon and in French. At the same time, *-ing* was very productive on borrowed French bases (e.g. *spusing* - ‘marriage’, *with ioiinge* ‘with rejoicing’) (Miller, 2002: 319).

In addition to the French influence (particularly on officialdom, at court and for trade), the emergence of verbal gerunds might have been marked by Latin influence (literature and religion). The grammatical teaching of English often relied on using classical Latin grammar as a model for English grammar. The frequent use of deverbal *-ing* forms as syntactic calques for Latin gerundials in English translations contributed to the productivity of English gerunds and facilitated the generalization of their functions (cf. example [20] and Miller, 2002: 369).



(20) In quo corrigit adolescentior viam suam? in custodiendo sermones tuos. (Vulgate Bible, Psalms 118:9)

In what þynge amendeþ þe ʒenge man his waie in keping þy wordes?  
(c1350, Helsinki corpus<sup>4</sup>, example taken from de Smet, 2008)

By what does a young man correct his way? By observing your words.

English at the same time was a vernacular language and survived under both of these influences as the language spoken by the common person. It was generally not learnt through schooling as the two above but as the mother tongue. This made it flexible and changing and allowed it to absorb all the influences.

After 1250, *-ing* had mainly replaced the *-ung* suffix. This original noun-forming use of this suffix was limited throughout the OE period to nouns of action (OED, 1971) and the use of this suffix to create gerund forms of any verbs (except from *may*, *shall* and a few other auxiliary verbs) only appeared in the early Middle English period. As the gerund was considered to be a noun it took objects such genitive objects. The use of accusative objects with gerunds was rare in OE and became more frequent in Middle English. Today, a gerund still can take a genitive object. However,

---

<sup>4</sup> The Helsinki corpus is a corpus of approximately 1.6 million words of English dating from the early Old English period to the end of the Early Modern English period.

when its verbal force is strong, it seems to prefer an accusative / direct object (Curme, 1980: 483).

Historically, gerunds and present participles did not always share the same morphological form. The English present participle was originally used like an adjective, which until the 14th century had different endings. In OE, present participles used the suffix *-ende*, which had been changed since the beginning of the ME period into *-inde* / *-ynde* (Southern Dialect), *-ende* / *-unde* (Midland Dialect) and *-ande* (Northern Dialect). That is, the present participle suffix had a completely different etymology from the gerund suffix. It began to appear in manuscripts spelled *-inge* only in the middle of the 14th century.

(21) Jhon was in desert baptisyng and preching...

(Wyclif, Mark 1:4, example taken from Cho, 1985)

John appeared in the desert, baptising and preaching...

Anglo-Norman scribes started confusing its written and spoken form (by then, generally *-inge*) with that of the gerund suffix (*-inge*) (Broderick, 2000). The verbal noun and the participle thus became merged into one form. As a result, the gerund began to develop verbal characteristics, probably under the influence of the participle. It developed tense distinctions and the passive voice preserving still its syntactic characteristics of a noun. Both the gerund and the present participle have acquired more verbal force

over time. In Late Modern English, the present participle has the same form like the gerund and can also take accusative objects.

Furthermore, there was some influence of the functions of the infinitive on the gerund in OE. Often, the infinitive ending *-n* and the gerund ending in *-ing* could probably not be clearly distinguished phonologically. This might have caused a certain confusion of participles and infinitives. It can be assumed that the primary origin of the gerund is based on *-ung* / *-ing* in OE and its secondary origin is based on the function of the infinitive in OE (Poutsma, 1929).

Intermediate gerund forms which are found in earlier stages of English and which would seem ungrammatical nowadays seem to provide evidence for the gradual transition from verbal nouns to gerunds. In (22) and (23) the *-ing* form still retains the determiner, but the preposition has already disappeared.

(22) Nothing in his life became him like the leaving it.  
(Shakespeare 1606, Macbeth Act 1, scene 4, 7–8)

(23) I was overtaken by the judgment of Heaven for my  
wicked leaving my father's house. (Defoe 1719,  
Robinson Crusoe, Chapter 1)

The origin of the gerund is rather complex and this section is only a condensed account of opinions expressed on this topic in the literature.

## 3 A typology of *-ing* forms

### 3.1 The *-ing* suffix in common nouns

English has some distinct morphemes sharing the phonological shape *-ing*. Traditionally grammarians divide *-ing* forms mainly into nouns, present participles and gerunds. The Oxford English Dictionary – not necessarily a reference with a strong focus on a detailed classification and morphological features of affixes, but often quoted as a standard authority on the English language – lists the following three *-ing* suffixes:

- ing 1: a derivational suffix
- ing 2: the present participle suffix
- ing 3: the gerund suffix

Fig. 2: *-ing* suffixes according to the Compact OED, 1971

For some people, there is a difference in the pronunciation of the stem according to what type of *-ing* is affixed so that words in which the *-ing* affix is a participial ending have one more syllable with respect to those where *-ing* is a nominalising suffix.

- (24) stars twinkling in the sky
- (25) in the twink/ə/ling of an eye

As a derivational suffix, *-ing* functions as a productive nominaliser. It can be attached to verbs to form a corresponding concrete or abstract noun (26) or compound (27). It occurs in words like *sibling*, *shilling* or *sterling*<sup>5</sup> where it also may represent the diminutive suffix *-ling*. Some gerunds of frequently used verbs seem to have established themselves as true nouns or seem to be on the verge of becoming nouns (cf. 6.4.2.2).

(26) ceiling, wedding, farthing<sup>6</sup>

(27) fishing boat, drawing book

The distinction between action nominals / gerundial nouns and gerunds and also between participle clauses and gerunds represents a particular challenge. Quirk et al. (1985, 1290f.) have tried to classify *-ing* forms according to a gradience from purely nominal to highly verbal (e.g. *The painting hang in the room.* → *I don't like him painting the house.*) There is a similar cline going from purely adjectival to purely verbal (e.g. *the interesting book* → *the sleeping man / the man sleeping* → *He is painting the girl.*)

---

<sup>5</sup> sibling: from O.E sabb "kinship", shilling: from \* skell- "to resound, to ring," or \*skel- "to split, to divide", sterling: probably from M.E. sterre "star", because of the stars on certain Norman coins

\*: not attested in any written source, but reconstructed by etymological analysis, <http://www.etymonline.com> (last accessed 8 August 2010)

<sup>6</sup> an old British coin

### 3.2 The *-ing* suffix in present participles

The present participle is used in the following contexts:

➤ **present-participial adjective / adjective complement**

(28) a welcoming atmosphere

(29) he was very welcoming

Present participles can function as adjectives modifying nouns or pronouns. Some participles seem to have become syntactically reanalysed and established themselves as true adjectives. This is indicated by the use of modifiers like adverbs such as *very* / *utterly* and comparative forms like *more* and *most* (followed by *than*), *not as ... as*, *more and more...* Others seem to fall in-between verbs and adjectives (cf. 32).

(30) a) the *boring* film

b) I found the film to be *boring*.

c) I found the film to be even more *boring* than the book.

(31) a) the *screaming* child

b) I don't want the child to be *screaming*. (no adjective, but progressive form!)

c) scarves in more *screaming* colours (stative verb)

d) \* this child is more *screaming* (ungrammatical with

dynamic verb)<sup>7</sup>

(32) a more and more *changing* society (adjective/verb?)

➤ **as equivalent to a preposition**

(33) I'll contact you *regarding* the contract.

(34) It was not possible *owing* to a lack of cooperation.

➤ **abridged, non-finite relative clause / apposition**

(35) the boy *cutting* flowers

(36) his current research, *investigating* the effects of...

Such participial phrases can be ambiguous between shortened relative clauses / appositions and adverbial adjunct clauses (37).

(37) John, *knowing* about their plans, didn't answer.

→ John, who knew about their plans ...

→ As / Although John knew about their plans...

➤ **adverbial adjunct clauses**

As a participle in adjunct clauses, the *-ing* forms may acquire adverbial meaning even when it is not preceded by a conjunction. The introductory phrase is set off with commas. Usually an

---

<sup>7</sup> \*: 'ungrammatical' / ?: 'dubious' or 'marginal'

introductory abridged clause has no overt subject. Its understood subject is therefore the same as the one in the main clause. Introductory clauses with an overt subject are rare (cf. 41).

#### CAUSE

- (38) *Seeing* their uneasiness, he observed them more attentively.
- (39) *Knowing* he could not win, he was scared.
- (40) a) *Having* found it, I hurried back home.  
b) cf. \*/?*Having* found it, it's far easier to use.
- (41) We didn't get home until midnight because of *the train not leaving* until 9 o'clock.

#### TIME

- (42) There are many ways to kill time, while *waiting* for your train.
- (43) *Having* said that, I still had a difficult choice to make.

#### MANNER

- (44) *Smiling* at each other, they left together.
- (45) They went home, *watching* the moon rise.

#### CONCESSION

- (46) Although *feeling* unwell, he went out.



CONDITION

- (47) *Sitting* on the couch watching TV, you will not lose weight.

➤ **progressive aspect**

The progressive form is often regarded as a distinct grammatical category with special functions. Is it preceded by a form of 'BE'. However, in abridged relative clauses and adverbial adjunct clauses the present participle seems quite similar to the progressive form and in those contexts the participle can also be interpreted as an abridged progressive form (cf. 49 & 50).

- (48) He is washing the car.  
(49) He was shot by a man (who was) hiding on the roof.  
(50) He broke down one day, (when he was) walking on a London street, tears (were) streaming down his face.

### 3.3 The *-ing* suffix in gerunds

The distinction between the present participle and the gerund is an issue which is often raised and debated by grammarians. The traditional view is that the gerund is a verb with substantival value, a verb form with a noun-like role in the sentence retaining characteristics of both verbs and nouns. The present participle is regarded to be a verb with adjectival or adverbial value. It cannot be preceded by possessive articles or possessive noun phrases. Both gerunds and participles are formed by adding *-ing* to a verb and both express action or a state of being. However their function within the sentence differs.

The gerund is often triggered by certain expressions such as:

- verbs / verb + preposition

avoid, deny, enjoy, finish

complain about, prevent from, look forward to, succeed in,  
insist on

- adjectives + preposition

tired of, interested in, keen on, fond of

- nouns / nouns + preposition

have trouble doing sth.

have difficulty (in) doing sth.

- prepositions

instead of, apart from

- idiomatic phrases

can't help doing, there's no point in doing, be used to doing

Fig. 3: Expressions followed by gerunds

These expressions need to be learned individually and often there does not seem to be a grammatical rule for the preference of gerunds over the infinitive. Some verbs take both forms, gerunds and infinitives, interchangeably; sometimes there is a difference in meaning (*stop, remember, try, forget*).

- (51) a) He stopped looking up and down the street. (→ He did not look any longer)  
b) cf. He stopped to (→ in order to) look up and down the street.
- (52) a) I remembered filling out the form. (→ that I had filled out the form)  
b) I remembered to fill out the form. (→ that I had to fill out the form)

Gerunds frequently occur after prepositions, *in, for, of, by, to* and *without* being the six most common prepositions combining with gerunds (De Smet, 2008). Additionally, gerunds can follow

the prepositions *instead of, before, after, on, through, from, besides, in fear of, for the sake of, on the verge of, except for* etc.

Gerunds, which can be single words or gerund phrases, are often said to have the same functions as NPs:

➤ **Subject gerunds**

- |      |                                                                                                              |
|------|--------------------------------------------------------------------------------------------------------------|
| (53) | <i>Talking</i> is essential.                                                                                 |
| (54) | <i>Working</i> in the US can be harder than expected.                                                        |
| (55) | It is no use him <i>pretending</i> otherwise. (gerund construction with a subject of its own in NP position) |

➤ **Object gerunds:**

- |      |                                             |
|------|---------------------------------------------|
| (56) | John prefers <i>walking</i> .               |
| (57) | She likes <i>watching</i> the light change. |
| (58) | We had listened to him <i>speaking</i> .    |

➤ **Gerunds as objects of prepositions**

- |      |                                                                  |
|------|------------------------------------------------------------------|
| (59) | They established a commission on <i>preventing</i> conflict.     |
| (60) | He was applauded for <i>tackling</i> such a controversial topic. |

## 4 A more detailed description of English gerunds

### 4.1 Gerunds as a mixed category

Gerunds are derived from verbs and act similar to nouns. Unlike participles, they cannot be formed by adding any other suffix to the verb stem. When they are not the object of a preposition, they often can be seen as interchangeable with the infinitive, although, as mentioned above, some verbs take only infinitives and others gerunds.

Gerunds can form a plural (61) or a genitive (mainly before *sake* (62), Jespersen, 1993: 261).

- |      |                                      |
|------|--------------------------------------|
| (61) | his <i>comings</i> and <i>goings</i> |
| (62) | reading for <i>reading's</i> sake    |

Being like nouns, gerunds can take adjectives:

- |      |                                                         |
|------|---------------------------------------------------------|
| (63) | <i>Brisk walking</i> is recommended for younger people. |
|------|---------------------------------------------------------|

Being like verbs, they can take adverbs:

- |      |                                                        |
|------|--------------------------------------------------------|
| (64) | <i>Walking slowly</i> is recommended for older people. |
|------|--------------------------------------------------------|

Several grammar books seem to evade a clear distinction between present participles and gerunds and they do look indeed quite similar, especially in adjunct clauses (Königs, 2004: 150).

- (65) Because of the buses not *running* (gerund / participle?),  
I was late for work.
- (66) After *talking* (conjunction + participle / preposition +  
gerund?) to you, I always feel better.

One could argue that gerunds can be distinguished from participles according to their grammatical function in a sentence e.g. subject or object position. It has also sometimes been said that gerund clauses cannot be omitted without making a sentence ungrammatical.

In the Cambridge Grammar of the English Language, the view is taken that a distinction between gerund and present participle cannot be sustained (Huddleston & Pullum, 2002: 82). Huddleston and Pullum only identify one single form, the gerund-participle, and attempt to subsume traditional gerunds, present participles and progressive forms under one category. Broderick (2000) also tries to establish a monosyntactic and monosemic analysis of *-ing* forms characterising all such forms with a common feature such as “action” and called it the “action inflectional suffix”. Historically, gerunds and present participles are derived from different sources, as discussed in the previous chapter.

We can observe some similarities between noun phrase (NP) and sentences (S) in general. Both sentences and noun phrases occur as subjects or direct objects and undergo passive (Abney, 1987: 23, e.g. ‘*John* was known by many linguists.’ vs. ‘*That John*

*came* was known by many linguists.’, cf. also Lees, 1960). The construction in which the noun phrase looks most like a sentence is the gerund (Abney, 1987:105).

According to Ross (1973), there is a range of structures having both sentence and noun phrase properties. These constructions form a continuum of “nouniness” with tensed S and concrete nouns at the endpoints of the spectrum:

**The scale of nouniness**

tensed S → indirect question → infinitive → Acc-ing → Poss-ing  
→ Ing-of → derived nominal → concrete noun

Fig. 4: The scale of nouniness according to Ross

At one endpoint of the spectrum, we find true nouns, which may contain determiners, be pluralised and take adjectival and not adverbial modification. At the other end, there are clauses that do not take determiners, cannot be pluralised and take adverbial modification only. In the middle, there are various gerund constructions, which may come with determiners, while others prohibit them. Some can take adjectival modification while others only allow adverbial modification. If we want to draw a clear cut between sentence and NP, it is probably between Acc-ing (the most NP-like sentence) and Poss-ing (the most sentence-like NP) (Reuland, 1983).

When trying to categorise *-ing* forms, it becomes obvious that part-of-speech categorisation can be a challenging task for linguists. *-ing* forms like gerunds could either be analysed as a verb or a noun and are generally viewed as mixed categories, forms that have both nominal and verbal features (Hudson, 2003: 611). It has been assumed that the *-ing* suffix is responsible for the external nominal properties while the verb stem is responsible for the internal verbal properties (Baker, 1985). The present participle, on the other hand, is often said to have only verbal features. The picture is more complex though as present participles can partly be reanalysed as prepositions or adjectives.

It is generally debatable whether a word can belong to two categories at the same time and whether there are fuzzy part-of-speech categories. Prescriptivists like Fowler are more uneasy with dual or multiple category membership. In his “King’s English” (1908), Fowler rejects the “fused participle” treatment of gerunds – i.e. that these forms are gerunds / nouns and participles / verbs at the same time. In traditional semantics, which involves the Aristotelian model of necessary and sufficient conditions, researchers believe that category membership depends on a fixed set of conditions or features. As a consequence, category membership is a binary and a clear-cut yes-or-no issue. This classic approach is challenged by the prototype theory of concepts developed by the psychologist Eleanor Rosch (1973). She pointed out that categories are defined by prototypes, examples of a category that come to mind first and represent the most typical



cases. This offers a different view of categorisation and implies fuzzy boundaries and graded membership. We will follow this approach and assume that there are mixed categories in grammar due to the historical development of languages. Constructions with mixed syntactic properties combine the external distribution of one category with the internal structure of another. Properties of mixed categories depend in a predictable way on their historical source, which somehow supersedes the discussion about a clear distinction between present participles and gerunds.

## 4.2 The structure of English gerunds (Poss-ing, Acc-ing, PRO-ing, Ing-of)

The following section describes the main different gerund constructions. There are a number of distinct structures in which the gerund appears (Abney, 1987)

### ➤ Poss-ing or full gerund

(67) his *singing* the song

The form which is traditionally called the full gerund has a subject in the Poss(essive) case. We shall refer to it as the Poss-ing construction. This construction combines properties of both noun phrases (NP) and sentences (S). In terms of its external distribution, it behaves like a NP and appears in NP positions such as the subject or object position. On the other hand, the remainder of the gerund constitutes a verb phrase (VP). Any verb (except from *may*, *shall* and a few other auxiliary verbs) can take the affix *-ing* and appear in this gerundive construction. In this regard, gerunds differ from usual derived nouns.

(68) Did his *reading* the poem upset you?

(69) Have you heard of John's *building* a navy?

The subject of the gerund receives genitive case and not nominative case and therefore behaves like a possessor of a NP.

- |      |                                        |
|------|----------------------------------------|
| (70) | John's <i>approving</i> this procedure |
| (71) | cf. John's approval of this procedure  |

➤ **Acc-ing or half gerund**

- |      |                                            |
|------|--------------------------------------------|
| (72) | I disapprove of him <i>smoking</i> cigars. |
|------|--------------------------------------------|

A second gerundive form, whose subject is in the Acc(usative) is referred to as the Acc-ing construction.

➤ **PRO-ing**

- |      |                                              |
|------|----------------------------------------------|
| (73) | I disapprove of (PRO) <i>smoking</i> cigars. |
| (74) | I avoided (PRO) <i>meeting</i> him.          |

In several studies on gerunds, we find a special type for subjectless gerunds, which may be interpreted like subjectless infinitives (Abney 1987: 168). It has been assumed that an empty pronominal (PRO) represents the gerund's subject. We will refer to this construction as PRO-ing. There are discordant views whether PRO-ing has a distinct structure or not. In the following, we will subsume PRO-ing under the Poss-ing, Acc-ing and Ing-of

categories and hold the view that PRO-ing patterns with these structures depending on the context.

➤ **Ing-of / gerundial noun / deverbial noun**

(75) After their *eating* of the fruit they were banned.

(76) I disapprove of his *smoking* of cigars.

There is an *-ing* deverbial noun, a form that lacks most verbal characteristics. This form is also known as the Ing-of construction or the gerundial noun.

### 4.3 Analysis of Acc-ing and Poss-ing gerunds as compared to Ing-of constructions

In this section, we shall have a closer look at the characteristics of gerunds, with special attention to Acc-Ing and Poss-ing gerunds as compared to gerundial nouns. Both Poss-ing and Acc-ing gerunds are clause-like with regard to their case-assignment. They take ordinary NP objects and PP<sup>8</sup> objects as complements and can be modified by (VP)<sup>9</sup>-adverbs.

(77) John's/John quickly leaving surprised everybody.

Poss-ing and Acc-ing can directly select for a complement, without need for 'of'.

(78) Mary's/Mary revising the book.

Both constructions are negated with 'not'.

(79) His / him not having left yet could be a sign of his reluctance.

---

<sup>8</sup> PP – prepositional phrase

<sup>9</sup> VP-adverbs have scope over VPs (unlike S-adverbs that have scope over entire sentences)

Nevertheless, we find some exceptions to this rule with ‘*no*’ or ‘*any*’ in prohibitions or existentials.

- |      |                                                    |
|------|----------------------------------------------------|
| (80) | No playing loud music!                             |
| (81) | No eating sweets in lectures!                      |
| (82) | There’s no denying it.                             |
| (83) | There isn’t any telling what he may say next week. |

There are several factors, generally influencing the choice between Acc-ing and Poss-ing<sup>10</sup>. Firstly, there can be a difference in meaning.

- |      |          |                                      |
|------|----------|--------------------------------------|
| (84) | Poss-ing | I dislike his singing in the movies. |
| (85) | Acc-ing  | I dislike him singing in the movies. |

(84) can be interpreted so that his singing as such is not appreciated whereas Acc-ing here suggests that the whole act of him singing is considered to be annoying.

Moreover, Poss-ing is usually considered to be more formal than Acc-ing.

- |      |                                                                   |
|------|-------------------------------------------------------------------|
| (86) | Would you mind John’s opening the window? [formal]                |
| (87) | Would you mind John opening the window?<br>[slightly less formal] |

---

<sup>10</sup> cf. <http://www.ling.cam.ac.uk/li8/gerunds.pdf> for Poss-ing / Acc-ing characteristics (last accessed 8 August 2010)

The subject position favours Poss-ing as compared to object position. Longer subjects however tend to favour Acc-ing in formal style. Some types of subject disallow or disfavour Poss-ing.

- |      |                                                    |
|------|----------------------------------------------------|
| (88) | I hate it (*its) raining when I'm walking home.    |
| (89) | I hate this (*this's) happening.                   |
| (90) | I hate there (*there's) always being some problem. |

Matrix verbs influence the choice of one form over the other.

- |      |                  |                          |
|------|------------------|--------------------------|
| (91) | appreciate, mind | favourable to Poss-ing   |
| (92) | stop             | unfavourable to Poss-ing |

The main difference between Poss-ing and Acc-ing is that the former patterns more with noun phrases while the Acc-ing is more similar to sentences. One example to demonstrate this is agreement. Poss-ing forms trigger agreement while Acc-ing forms pattern with sentences (Abney, 1987: 111).

- |      |          |                                                                            |
|------|----------|----------------------------------------------------------------------------|
| (93) | S        | That John came and that Mary left bothers / (*bother) me.                  |
| (94) | Acc-ing  | John coming (so often) and Mary leaving (so often) bothers / (*bother) me. |
| (95) | Poss-ing | John's coming and Mary's leaving bother me.                                |
| (96) | NP       | Bill's arrival and Mary's departure bother me.                             |

Acc-ing and Poss-ing gerunds also show differences with regard to long-distance binding of their subjects.

- |      |          |                                                                    |
|------|----------|--------------------------------------------------------------------|
| (97) | Poss-ing | They thought that each other's giving up the ship was forgivable.  |
| (98) | Acc-ing  | ?/*They thought that each other giving up the ship was forgivable. |

Likewise, (99) and (100) show that Acc-ing and Poss-ing gerunds differ in their syntactic behaviour.

- |       |          |                                                                                         |
|-------|----------|-----------------------------------------------------------------------------------------|
| (99)  | Acc-ing  | a) We remember him describing Rome.<br>b) cf. the city that we remember him describing  |
| (100) | Poss-ing | a) We remember his describing Rome.<br>b) cf. *the city that we remember his describing |

Poss-ing gerunds containing wh-subjects can front under pied-piping, not so for Acc-ing gerunds.

- |       |          |                                                                                |
|-------|----------|--------------------------------------------------------------------------------|
| (101) | Poss-ing | the pianist [whose playing with the orchestra last season] was so electrifying |
| (102) | Acc-ing  | *the pianist [who playing with the orchestra last season] was so electrifying  |



Moreover, the subject of Poss-ing gerunds can take wide scope while that of Acc-ing strongly prefers narrow scope.

- |       |          |                                                               |
|-------|----------|---------------------------------------------------------------|
| (103) | Poss-ing | John disapproves of everyone's taking a day off. (wide scope) |
| (104) | Acc-ing  | John disapproves of everyone taking a day off. (narrow scope) |

Other sentential aspects of Poss-ing are that it prefers adverbs over adjectives and can take double object complements.

- |       |                              |
|-------|------------------------------|
| (105) | despite her giving him a hug |
|-------|------------------------------|

The Ing-of construction differs syntactically from both Acc-ing and Poss-ing. It can occur as a bare nominal or accompanied by a determiner such as the definite article or a pronoun. It behaves like other any deverbal noun (*arrival, composition* etc.). Nevertheless, most Ing-of constructions are no distinct lexemes, in contrast to common nouns ending in *-ing*, or have not yet established themselves completely as lexemes.

This construction does not allow any auxiliaries of aspect and voice (106 & 107) and triggers plural agreement (108). Furthermore it is negated by negative pronouns or the negative determiner 'no' (109). The object of the verbal noun receives the analytic genitive case, marked by the preposition 'of'. The subject of a verbal noun gets the synthetic genitive case (e.g. 110).

Usually, a determiner precedes Ing-of when it has no subject.  
Adverbial modifiers are replaced by corresponding adjectives.

- |       |                                                         |
|-------|---------------------------------------------------------|
| (106) | *the Queen's having opened of the new building          |
| (107) | *the Queen's having been opening of the new building    |
| (108) | all openings of new buildings                           |
| (109) | no checking of the staff is ever made                   |
| (110) | the Queen's opening of the new building                 |
| (111) | the Queen's unexpected(*ly) opening of the new building |

## 5 German translation equivalents for English gerunds

The gerund is a peculiarity of the English language that does not entirely correspond to gerund-like forms in other languages which were discussed in Section 2.1. As mentioned above, gerund-like forms in Romance language, for instance, are more similar to English participles. German itself has not developed a gerund or a gerund-like form. The German present participle, which resembles the English gerund only with regard to very few aspects, is generally used rather differently in German. There are other grammatical structures in German that translators can typically choose to translate clauses containing English gerunds. These are basically infinitives, nominalised infinitives and deverbal nouns (particularly those ending in the suffix *-ung*).<sup>11</sup>

English participle clauses, on the other hand, may often be translated by using German participles, although this option is not always possible. In general, German seems to prefer infinitives and periphrastic constructions with prepositions or adverbial sub-clauses instead. German participles however can easily be converted into nouns:

---

<sup>11</sup> Similarly to German, Dutch has not developed a gerund either and can also use nominalised infinitives instead (e.g. ‘Op vakantie gaan is leuk.’ - Going on vacation is nice.) Additionally, *-ing* in Dutch is a derivative ending that can be added to a restricted number of verbs to form a feminine verbal noun like in German (e.g. *zegening* / *Segnung* – blessing; *verdediging* / *Verteidigung* – defense; *verbetering* / *Verbesserung* – improvement).

(112) present participle: *sprechend* (speaking)

→ nominalisation: *der Sprechende* – the speaking person

(113) present participle: *gekocht* (cooked)

→ nominalisation: *das Gekochte* – the cooked food

Furthermore, several German prepositions combine easily with infinitives. This seems to support the theory that gerunds emerged in English to fill a syntactic gap as English prepositions do not combine with infinitives:

(114) *ohne zu wissen* – without knowing

(115) *anstelle zu klagen* – instead of complaining

Translators often have to compensate for the fact that both English participles and gerunds are potentially ambiguous without complements. Translating *-ing* forms often involves obligatory or optional grammatical shifts. Translators may have individual criteria for choosing a certain translation or choosing from a set of possible translations, but these criteria also depend on the register and stylistic level of a text. There are often similar translation possibilities for translating either gerunds or present participles into German (cf. also Königs, 2004):

➤ **participial phrase functioning as an adverbial clause**<sup>12</sup>

- (116) although *having been warned* - obgleich gewarnt  
(117) *reaching* the river – den Fluss erreichend

➤ **(explicit) adverbial sub-clauses**

- (118) *Having finished* his work, he played computer.  
Nachdem er seine Arbeit beendet hatte, spielte er  
Computer.
- (119) He has been detained for two months without *being brought* before the court.  
Er wurde für zwei Monate festgehalten, ohne dass er vor  
Gericht gebracht wurde.

➤ **sub-clause with inversion**

- (120) *If being followed* by a car,  
Würde ich von einem Auto verfolgt, so ...

---

<sup>12</sup> This is not possible with all verbs and it sounds rather formal and should not be used in all text types.

➤ **main clause (semantically implicit)**

(121) He has been detained for two months without *being brought* before the court.

Er wurde für zwei Monate festgehalten und (er) wurde nicht vor Gericht gebracht.

(122) After *having finished* the book, I got up.

Ich hatte das Buch zu Ende gelesen. Ich stand auf.

➤ **Inflected verb**

(123) They agree in *describing* their goal...

Sie beschreiben ihr Ziel übereinstimmend...

➤ **infinitives/ infinitive clause**

(124) He has been detained for two months without *being brought* before the court.

Er wurde für zwei Monate festgehalten, ohne vor Gericht gebracht zu werden.

(125) They talked about *quitting*.

Sie haben darüber gesprochen zu kündigen.

➤ **nominalised infinitive / compound**

(126) *Swimming* is an ideal workout. - Schwimmen ist ein idealer Ausgleichssport.

(127) *Wasting* time is over. - Das Zeitvergeuden hat ein Ende.

➤ **(deverbal) noun (+preposition)**

(128) after *finishing* his work - nach Beendigung seiner Arbeit

(129) *having been warned* - trotz Warnung

(130) by *mentoring* the Afghan police - durch Mentoring<sup>13</sup> für die afghanische Polizei

➤ **noun/verb combinations (Funktionsverbgefüge)**

(131) *assisting* developing countries  
Entwicklungsländern Beistand geben

(132) *considering* the options  
die verschiedenen Möglichkeiten in Betracht ziehen

---

<sup>13</sup> German has borrowed *ing-forms* from several English verbs reanalysing them as nouns, e.g. *Mentoring, Camping, Jogging, Marketing*

➤ **adverbs** (often when translating phrases such as *start by, begin with, succeed in* and *the necessity of*)

- |       |                                   |                                     |
|-------|-----------------------------------|-------------------------------------|
| (133) | we can start by <i>saying</i> ... | wir können zunächst einmal sagen... |
| (134) | they continue <i>selling</i> ...  | sie verkaufen weiterhin...          |

➤ **adjectives**

- |       |                                          |                                    |
|-------|------------------------------------------|------------------------------------|
| (135) | by <i>improving</i> the vaccine          | durch einen verbesserten Impfstoff |
| (136) | <i>enlarging</i> the powers of the court | größerer Einfluss des Gerichts     |

➤ **prepositional phrase**

- |       |                                     |
|-------|-------------------------------------|
| (137) | <i>using</i> an axe – mit einer Axt |
|-------|-------------------------------------|

➤ **change in perspective / passivisation**

- |       |                                              |                                               |
|-------|----------------------------------------------|-----------------------------------------------|
| (138) | by <i>exposing</i> firms to more competition | Firmen werden stärkerem Wettbewerb ausgesetzt |
| (139) | by <i>negotiating</i> agreements             | Abkommen werden ausgehandelt                  |



➤ **ZERO equivalent**

(140) after *finishing* his work - nach (ZERO) der Arbeit

(141) He put off *making* a decision – Er schob die  
Entscheidung (ZERO) auf.

Often several English lexemes may be amalgamated into one word in German.

(142) it is worth noting, it is worth pointing out  
es ist bemerkenswert

## **6 The analysis of gerunds in electronic corpus data**

### **6.1 Corpus linguistics and corpus-based translation studies**

Corpus linguistics is a method in different branches of linguistics that uses language corpora as samples of the language and as a source of data and evidence in linguistic enquiry. Corpora are used for instance in computational linguistics, phonology, historical linguistics, lexicology and lexicography, language teaching and translation research and practice.

Working with a large corpus requires the use of advanced technology. A corpus is a finite-sized body of machine-readable text which was selected, chosen or assembled according to explicit criteria and sampled in order to be maximally representative of the language variety / genre / register under consideration (Mc Enery & Wilson, 2001: 32). Technological manufacturing advances have improved computing power and vastly reduced storage costs which, together with the shift from analogue to digital storage of text, voice and video, have made available large quantities of corpus data and the possibility to annotate, align and analyse these data with the aid of dedicated software programmes. These advances when combined have the potential to accelerate

quantitative data analysis and scientific investigations into linguistic issues considerably.

The use of corpora is widespread in linguistics but comes under criticism from some quarters, e.g. from Chomsky. He argues that finite corpora are insufficient to be adequately representative of infinite languages. In his opinion, corpora are artefacts, incomplete and skewed. Some sentences might never occur in corpora because they are obvious, false or impolite for instance (Chomsky, 1962: 159). In a recent interview, Chomsky reasserted his position: “Corpus linguistics doesn’t mean anything. It’s like saying suppose [...] physics and chemistry decide that instead of relying on experiments, what they’re going to do is take videotapes of everything that’s happening and from that maybe they’ll come up with some generalisations or insights. Well, you know, sciences don’t do this.” (Chomsky in Andor, 2004: 97). This thesis will not attempt to discuss the limitations and advantages of empiricism and rationalism or whether truly empirical linguistics is possible. Nevertheless, the view is taken that corpora are a useful tool in linguistics *in addition to*, not *instead of* other approaches.

There is a wide array of different corpus types: spoken corpora vs. written corpora, general vs. specific corpora and synchronic vs. diachronic corpora. Dynamic / monitor corpora are constantly updated, in order to track language change for example. Monolingual corpora contain texts in only one language. Comparable corpora are comprised of texts in different languages

and are similar in content and form. Parallel corpora contain texts in one language along with their translation in another. The last-named are particularly important for the use in statistical machine translation. Nowadays, the prevailing methodology in machine translation is the use of phrase-based statistical models involving machine-learning techniques. These can be combined with rule-based approaches. In phrase-based translation, source language sequences of words are mapped into phrases in the target language during the translation process of a sentence using a probabilistic phrase translation table.

To ensure valid results, several criteria have to be considered when designing a language corpus: A well-formulated sampling frame and an idea of the purpose of the corpus are crucial. The corpus size plays an important role. A corpus must be large enough to provide statistically significant results. Meta-data about where the corpus data come from have to be considered. Representativeness as “the extent to which a sample includes the full range of variability in a population” (Biber, 1993) is a major issue in corpus design and determines the generalisability of the results of the research. Furthermore, scientific inquiries in a corpus have to be experimentally reproducible.

The utility of a corpus is increased when it has been annotated with linguistic information. Annotated corpora are becoming ever more important in linguistic research. Annotation does not involve adding new information but making implicit linguistic information explicit. It is, however, always based on a certain interpretation of

the data. The annotation may, for instance, include information on word classes. In POS-tagged corpora, part-of-speech (POS) tags have automatically been assigned to each token. Additionally, corpora may be annotated with information on morphological components and the lemmas<sup>14</sup> of the words (stemming or lemmatisation). Syntactically annotated (parsed) corpora are called tree-banks. Corpora may also be semantically tagged, discourse tagged, pragmatic / stylistic tagged or problem-oriented tagged with automatic, semi-automatic or manual methods.

Different types of corpus queries exist: a Keyword-in-Context (KWIC) Concordance lists specified strings in a text corpus surrounded by the text they are embedded in. Other corpus queries may display and count certain patterns, word-and-POS pairs and can also include a search for attribute values, a start- or end-tag or a wildcard character which will match any element of a specific type. Complex queries may then be constructed combining these aspects.

---

<sup>14</sup> lemma: abstract form of a word representing all word forms belonging to the same word (e.g. *go*, *goes*, *going*, *went*, and *gone* are lemmatised to *go*)

## 6.2 The CroCo corpus

The CroCo corpus is a linguistically annotated and aligned comparable and parallel corpus of German and English. It contains English original texts and their German translations as well as German originals and their English translations.

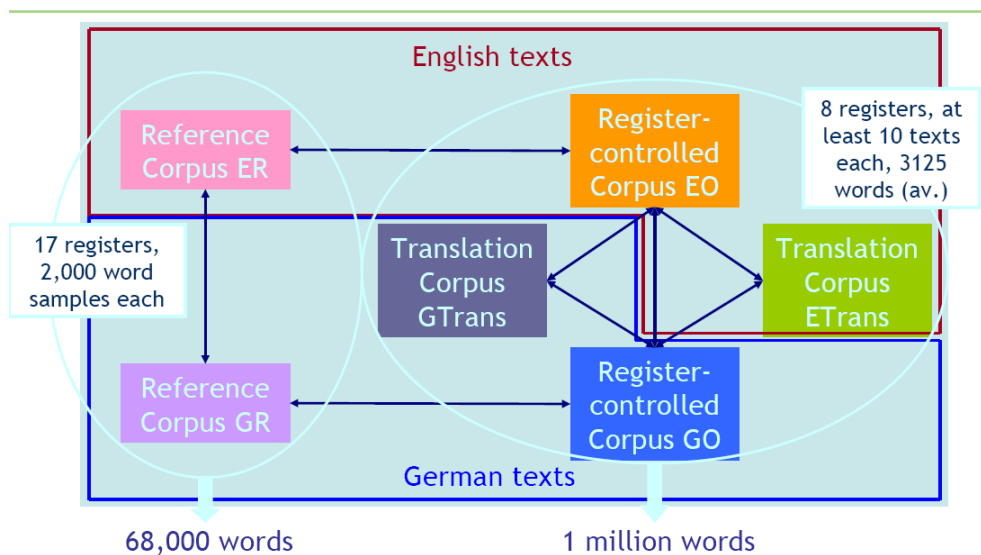


Fig. 5: The CroCo corpus, figure taken from Hansen-Schirra et al., 2006

This corpus was designed primarily for the investigation of translation shifts, particularly explicitation in translations for the English-German language pair.

Translations often contain lexical, grammatical, and semantic shifts. Lexical shift may occur when there is a gap in the lexicon of the target language (sometimes referred to as ‘lacuna’). Grammatical shifts, or transpositions, refer to changing

grammatical categories such as tense, number, person or part-of-speech. A shift of both lexis and grammar is called level shift (Catford, 1965: 73). Explicitation is a sub-class of semantic modifications in an annotation scheme for annotating translation shifts by Cyrus (2006). Explicitation can be a tendency or communication strategy of translators to include additional information in the translation that was not present on the surface structure in the source text but might have been inferred by the reader. Implication is the opposite and might also be a result of the translation process.

The CroCo corpus contains texts from eight written ‘registers’ (or rather text types): political essays (ESSAY), fictional texts (FICTION), instruction manuals (INSTR), popular scientific texts (POPSCI), letters to shareholders (SHARE), prepared speeches (SPEECH), tourism texts (TOU) and websites (WEB). Whether a text belongs to a certain register is determined by various parameters e.g. “field of discourse, “tenor of discourse” and “mode of discourse” (cf. Halliday and Hasan, 1989). There are at least ten texts per register with a length of circa 3,125 words (i.e. tokens without punctuation marks), a size that is considered sufficient (Biber, 1993). In total, the corpus contains approximately one million words. It includes two register-neutral reference corpora containing texts of 2,000 words from 17 registers in both languages. The reference corpora were compiled to level out register-specific peculiarities providing a wide range of registerial spread. The CroCo corpus is tokenised

and annotated for part-of-speech, phrase structure and grammatical functions. The author of this thesis participated in the annotation process. In the CroCo research project, every corpus text has been annotated with meta-information based on a detailed register analysis (cf. Fig. 6 and 7). Additionally, a sentence alignment between the subcorpora of original texts and translations has been conducted.



ESSAY	FICTION	INSTR	POPSCI	SHARE	SPEECH	TOU	WEB
FIELD							
EXPERIENTIAL DOMAIN							
foreign policy, economics vocabulary	general vocabulary	vocabulary of handling an object	specialised scientific vocabulary	business vocabulary	vocabulary on security policy	vocabulary of places of interest	mixed vocabulary
low variation of lexical items	highest variation of lexical items	lowest variation of lexical items	medium variation of lexical items	medium variation of lexical items	medium variation of lexical items	high variation of lexical items	medium variation of lexical items
medium variation of lexical verbs	highest variation of lexical verbs	lowest variation of lexical verbs	medium variation of lexical verbs	medium variation of lexical verbs	medium variation of lexical verbs	medium-high variation of lexical verbs	medium variation of lexical verbs
-	shorter lexical chains, spread over longer distances	-	longer lexical chains in fewer sentences	-	-	-	-
GOAL ORIENTATION							
Exposition	Narration	Instruction	Exposition	Exposition	Argumentation	Exposition	-
TENOR							
SOCIAL ROLE							
equal	equal	(higher)	equal	(lower)	equal	equal	equal
SOCIAL DISTANCE							
Consultative	Casual	Neutral	Neutral	Neutral	Neutral	Neutral	Consultative
MODE							
LANGUAGE ROLE							
highly constitutive	constitutive	less constitutive	constitutive	highly constitutive	highly constitutive	constitutive	constitutive
MEDIUM							
written	tendency towards spoken	written	written	written	written	written	written

ESSAY	FICTION	INSTR	POPSCI	SHARE	SPEECH	TOU	WEB
<b>FIELD OF DISCOURSE</b>							
<b>EXPERIENTIAL DOMAIN</b>							
Germany, economics vocabulary	general vocabulary	some procedural vocabulary	vocabulary discussing human behaviour	business, money vocabulary	Germany, Europe vocabulary	many verbs	mixed vocabulary
medium variation of lexical items	high variation of lexical items	lowest variation of lexical items	medium-low variation of lexical items	medium variation of lexical items	medium variation of lexical items	high variation of lexical items	medium variation of lexical items
medium variation of lexical verbs	high variation of lexical verbs	lowest variation of lexical verbs	high variation of lexical verbs	low variation of lexical verbs	medium variation of lexical verbs	medium variation of lexical verbs	high variation of lexical verbs
-	shorter lexical chains, spread over longer distances	-	longer lexical chains in fewer sentences	-	-	-	-
exposition	narration	instruction	argumentation	exposition	exposition	exposition	-
<b>GOAL ORIENTATION</b>							
<b>TENOR OF DISCOURSE</b>							
<b>SOCIAL ROLE</b>							
equal	equal	higher	equal	lower	equal	equal	equal
<b>SOCIAL DISTANCE</b>							
neutral	casual	consultative	neutral	neutral	consultative	neutral	consultative
<b>MODE OF DISCOURSE</b>							
<b>LANGUAGE ROLE</b>							
constitutive	constitutive	less constitutive	constitutive	constitutive	constitutive	highly constitutive	constitutive
<b>MEDIUM</b>							
written	tendency towards spoken	written	written	written	written	highly written	written

Fig. 7: Profiles of the German original registers (Neumann, 2008: 282)

### 6.3 Some previous corpus-based studies on gerunds

A few corpus-based studies have been carried out on different aspects of gerunds. Particularly historical aspects have been covered largely in diachronic comparisons. Fanego (1996, 2004) for example, examined the development and use of different gerund constructions throughout different stages of the English language on the basis of data from the Helsinki Corpus. Houston (1989) also investigated the gerundial usage from the 10th to the 17th century, Expósito (1996) looked at nominal gerunds in Chancery English and Kadeeri (2008) examined gerunds in Shakespearean English. Their respective research seems to confirm that English gerunds developed from abstract deverbal nouns that began to acquire more and more verbal properties. De Smet (2008) defines three main gerund categories that he looked for in the Helsinki Corpus: verbal gerunds (*eating the apple*), bare nominal gerunds (*eating of the apple*) and definite nominal gerunds (*the eating of the apple*).

This exemplifies the problem of a common classification of gerunds and the difficulty to build on previous research if another classification is adopted (e.g. according to Abney, 1987). De Smet's corpus research on gerunds has shown that the first instances of verbal gerunds can be dated back to the oldest period (1250–1350) of the corpus data examined from the Middle English period. Throughout the Middle and Early Modern

periods, gerund constructions were observed in the data with shifting frequencies of use. Their spread across the grammar of English extended over a period of several centuries. There was a dramatic rise of bare nominal gerunds halfway through the Middle English period and a rise of verbal gerunds which were syntactically more flexible. Subjectless gerunds (e.g. “*I avoided meeting Mary*”) spread at the expense of both *to*-infinitives and *that*-clauses. De Smet (2008: 69ff) claims that his corpus research leads to the conclusion that the development of gerunds was functionally motivated. Initially gerunds occurred almost exclusively after prepositions. Due to the absence of infinitives, gerunds substituted for the infinitive in prepositional environments and therefore filled a syntactic gap.

Also using the Helsinki Corpus, Núñez-Pertejo (1996) compared *-ing* forms of the type *a-hunting*, *keep the pot a-boiling*, to expressions which kept a full preposition: *be in hiding*, *be long in coming*, etc. *A*-phrases probably are derived from prepositional patterns of the type *on* + verbal noun in *-ing/-ung*, apparently in use from OE times. During the ME period, this preposition weakened to ‘*a*’ and was finally lost through aphesis. Additionally, Núñez-Pertejo (2001) conducted corpus-based studies on progressive *be* + *-ing* in Early Modern English and compared it to gerunds. Similar studies have been made by Smitterberg (2005).

Fewer studies have dealt with synchronic aspects of gerunds. One example is Lyne (2006, 2007) who used subsets of the British National Corpus to look for possessive determiners and objective pronouns preceding *-ing* forms (Fig. 8).

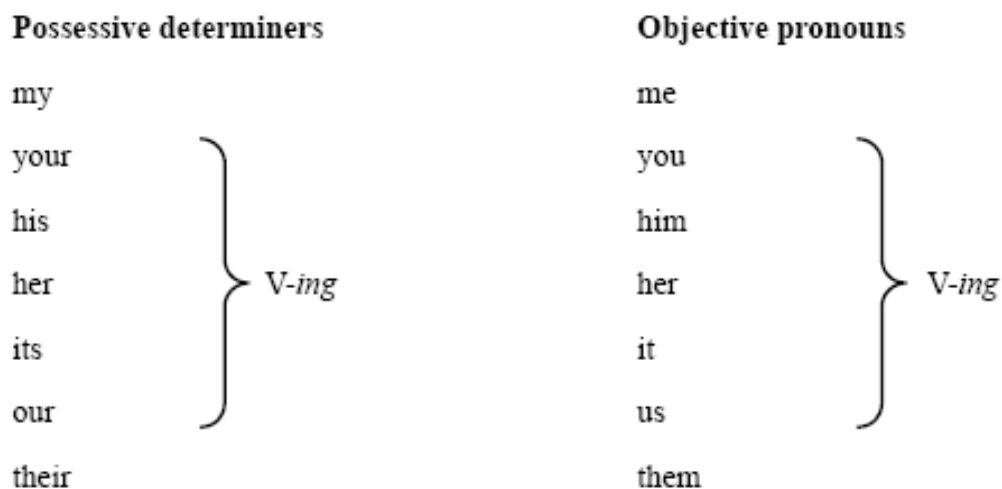


Fig. 8: Possessive determiners and objective pronouns + *V-ing*, Lyne, 2006

The subcorpora in Lyne’s study were divided into four text categories: Academic Prose, Fiction, Newspaper Texts and Spontaneous Conversation. A four-million-word subcorpus yielded a total of 300 verbal gerund clauses with pronominal subjects: 65 with the possessive form and 235 with the objective form. The genitive + gerund construction was mainly used in the Academic Prose category, followed by Fiction and News. Only one out of the 65 examples was found in Spoken Conversation.

Lyne's research on verbal gerund clauses with internal subject, and the variation between genitive and non-genitive forms of the subject seems to confirm that the genitive form before a gerund is very rare in present-day English. It can predominantly be found in formal registers. Common-case forms are seen to be spread more evenly across genres. Moreover, the linguistic factors phonology, animacy and NP length are seen to have influence on the choice of form. However, according to modern grammar books, the non-genitive forms are considered standard in contemporary language, with the genitive being regarded as a formal variant (e.g. Quirk et al, 1985: 1194, Huddleston and Pullum, 2002: 1192).

In her corpus queries, all types of *-ing* forms were included to make sure that no pronoun + verbal gerund was overlooked. This also led to a large number of irrelevant examples, which had to be deleted manually in a rather time-consuming process. These instances were mainly of the following types (cf. Lyne, 2006: 41):

(143)	trouble sort of thing, but er <pause> I don't know whether they bought <u>it thinking</u> they'll have all this money or what but er	KD8 6601 (conv.)
(144)	I can see <u>him playing</u> an anchor role in midfield.	K4S 1172 (news)
(145)	He was still in <u>his riding</u> clothes, well-cut jodhpurs and an old tweed jacket.	J54 1035 (fict.)
(146)	In Dr Johnson's all too familiar phrase, opera is an exotic and irrational entertainment. Is <u>it becoming</u> more irrational and exotic than it can handle?	AJF 113 (news)
(147)	But then <u>your implying</u> something aren't you	KD0 692 (conv.)

In (143) and (144), the *-ing* form is a present participle. Additionally, in (144), we have a construction that is typically found with certain complex-transitive verbs such as *see*, *hear*, *catch*, and *find*. These verbs take a direct object, in this case *him*, and *playing* is the object complement (Huddleston & Pullum, 2002: 1192). Riding in (145) is a part of the compound noun *riding clothes*. Many query hits had to be excluded because of inverted word order in questions such as (146) or due to the improper use of the possessive *your* instead of the contraction *you're* (147). This shows how difficult it is to find only relevant gerund forms via corpus queries.

Mehl (1995, 1996) investigated the translation of English gerunds into German to improve English-to-German machine translation (MT) systems. His studies are among the few that

focus on language contrastive aspects and the translation of gerunds into German. Mehl used Lyons' two volume reference book on semantics (1977) and its translation into German (1980, 1983) as a text corpus to find out how many gerunds are translated as common nouns, how many as nominalised infinitives, how many as verbs and on what conditions is one variant preferred over the others. The overall frequency distribution was as follows:

Verbs:	70 (65%)
Common nouns:	18 (17%)
Nominalised infinitives:	20 (18%)

Fig. 9: Frequencies of German translation options for gerunds in corpus data by Mehl (1996)

Mehl showed that syntactic transfer still represents a serious challenge in MT. Translation shifts and translation strategies depend on numerous syntactic, semantic and stylistic factors. As German has no gerund construction, gerunds are usually translated using subordinate clauses or nominalisations in MT. Both alternatives are not always possible or not equally fortunate depending on the following criteria (the following passage is taken from Mehl [1995]; his English examples were taken from Lyons, 1977, and the German translations are from the German translation of Lyons' book):



➤ **availability of a nominal derivative**

(148) Before embarking upon the discussion of this question,...  
Bevor wir uns der Diskussion dieser Frage widmen, ...

There is no German noun expressing the act of embarking (upon a discussion), hence the sentence has been translated subordinate clause with a finite verb. Here, the German subordinate clause requires a subject that is not explicitly mentioned in the English gerund clause.

➤ **ambiguity (in the respective context) and stylistic features of this noun**

(149) For example, 'Abiogenesis is spontaneous generation' can be understood as expressing, indirectly, a proposition about 'abiogenesis'.  
Zum Beispiel kann der Satz 'Urzeugung ist selbsttätiges Entstehen von Leben' so verstanden werden, dass mit ihm indirekt eine Proposition über das Wort 'Urzeugung' ausgedrückt werden soll.

The German noun *Ausdruck* denotes the act of expressing. However, this noun is ambiguous in a similar way as the English expression: it can mean an act (as in *Take these flowers as an expression of our gratitude.*) as well as an object (as in *This is not a well-formed logical expression.*) This ambiguity can be avoided easily by translating the gerund as a subordinate clause. Not every lexical ambiguity will enforce this decision: Readings that can easily be discarded in a particular context will not hinder the interpretation of an ambiguous word. Additionally, rare words (or readings) as well as words that do not fit the stylistic register of the text will decrease its readability.

➤ **possibility to combine the noun with all complements**

(150) The fact that the term 'expression' is in existence does not, of course, constitute sufficient reason for distinguishing it from 'lexeme', on the one hand, and from 'form', on the other.

Die Tatsache, dass der Terminus 'Ausdruck' existiert, stellt jedoch natürlich keinen genügenden Grund dafür dar, ihn von 'Lexem' einerseits und 'Form' andererseits zu unterscheiden.

In this example, one complement of the gerund verb consists of a pronoun. The respective German noun, however, does not favour the use of a pronoun complement (*?seine Unterscheidung von...*). Therefore, a subordinate clause is used in the translation.

➤ **syntactic and semantic ambiguity**

(151) Having made this point and given it due emphasis,... Nachdem wir dies festgestellt und gebührend betont haben...
------------------------------------------------------------------------------------------------------------------------------

(151) can be translated as: *Nach der Feststellung dieses Punktes und seiner gebührenden Betonung...* However, this implies an ambiguity that occurs systematically when a constituent in the German genitive might fit different semantic roles. In such cases, the subordinate clause variant makes relations more explicit.

➤ **stylistic evaluation of both alternatives**

Even if its interpretation is unambiguous, a noun phrase containing several modifiers and complements is hard to understand. In any case, constructions with multiply embedded noun phrases (which are easy to build in German) should be avoided. In the following example, a gerund verb with two complements, one of which is very complex, has actually been translated as a noun, yielding a clause in which head (*Einfügung*

[insertion]) and second modifier (*vor 'John' [before 'John']*) have been torn apart by the first modifier:

(152) [...] it can be made clear by inserting the phrase 'the name', or some similar descriptive expression, before 'John '.

[...] dies kann durch die Einfügung des Ausdrucks 'der Name', oder eines ähnlichen beschreibenden Ausdrucks vor 'John' klar werden.

Mehl came to the conclusion that most of the time the realisation of an English gerund as a German verb is the best alternative and often sounds more natural than a nominalisation. However, subordinate clauses may become confusing if too many of them are lined up or embedded into each other. A nominal realisation may lead to long distances between a subject and its predicate which are difficult to process for a human reader. He found nominalisations especially inappropriate in the case of reflexives or tenses other than the present tense:

(153) The room must be booked in advance by contacting Mr. Smith.

?Der Raum muss im Voraus gebucht werden durch Sich Wenden an Herrn Smith.

Mehl stated that the distinction between gerunds and participle constructions is generally a problem (e.g. *They discussed walking on the beach*). While even qualified human translators might have some difficulties in deducing the intended reading from the context and their world knowledge, these constructions pose a considerable challenge in MT. Gerund clauses often start with a preposition, which is not possible for participle clauses. Nevertheless, MT systems sometimes misinterpret the preposition as a particle, and the gerund as a participle introducing a relative clause.

Finding only relevant examples through corpus queries proved difficult in Mehl's study. He concluded that the automatic disambiguation of *-ing* forms is hardly possible due to the amount of encyclopaedic knowledge required. Often only the syntactic and semantic context proves which reading is intended and therefore, in his study, all types of *-ing* forms were collected via queries and disambiguated manually.

## **6.4 Gerunds and their translations the parallel corpus *CroCo***

### **6.4.1 Developing a query mechanism**

It is rather challenging to identify gerunds in a corpus for a number of reasons. Their mixed-category status between nouns and verbs and the lack of a commonly accepted definition among linguists present a formidable challenge. Previous corpus-based studies on gerunds focussed on different phenomena in corpora, based on their specific classification of gerunds. This is an obstacle for developing consistent query algorithms and building up on previous research. Even if there was a common understanding of what gerunds are, they occur in very different patterns and functions, triggered by fixed expressions, as arguments after verbs (that take either to-infinitives or gerunds, in some cases both), as gerunds clauses in subject or object function etc. This means that no corpus query would cover all gerund cases.

In several previous studies on gerunds, all *-ing* forms were queried so that no gerund could be overlooked. Automatic disambiguation of *-ing* forms proved too difficult. Therefore these forms had to be disambiguated manually (cf. Lyne, 2006; Mehl, 1995, 1996; de Smet, 2008). Manual disambiguation is relatively easy if the corpus is not particularly large (Mehl) or if only certain

phenomena are taken into account (e.g. gerunds after prepositions (de Smet), gerunds after pronouns (Lyne)).

Corpus queries in CroCo can provide information about frequency statistics of certain words or strings of characters. Finding all *-ing* forms thus poses no problem. Queries can also determine the context in which each word / string is centred (e.g. 10 words) and look for word co-occurrences (e.g. *-ing* forms preceded by pronouns). The query language is interpreted by the Corpus Query Processor (CQP, cf. Evert, 2005). Basically, queries in CroCo are used to locate certain sequences of characters, parts of speech and / or grammatical functions.

A first test query<sup>15</sup> in CroCo in the subcorpus of English originals of political essays (E2G\_ESSAY\_EN)<sup>16</sup> revealed the following POS-distribution of *-ing* forms:

---

<sup>15</sup> [word="\*.ing"], cf. ‘Susanne’ tagset for English (Sampson 1995, Evert, 2005)

<sup>16</sup> E2G: English to German, EO: English original, G2E: German to English, GO: German original

749	VVG	present participles
56	II	preposition
56	NN1	noun
42	JJ	adjectives
14	VBG	being
12	VDG	doing
12	VV0	base form
8	JB	adjective
8	PN1	anything, everything, nothing, something
5	NNL1	L+C noun
4	NNT1	singular time noun that can head a noun phrase functioning adverbially, NNT1h name of holiday or season
2	VHG	having
2	VVGK	going as catenative

Fig. 10: POS-distribution of *-ing* forms in E2G\_ESSAY\_EN

When ruling out prepositions, nouns, adjectives and pronouns automatically and looking only at *-ing* forms tagged as verbs<sup>17</sup>, the query results include also a few verbs ending in *-ing* (e.g. *bring* (E2G\_FICTION: 17123)) that are of no particular concern in a study about gerunds. Moreover, there are still too many irrelevant examples of other word categories that have been tagged wrongly

<sup>17</sup> [word=".\*ing" & pos="V.\*"]



(e.g. *Boeing* (EO\_SPEECH 23422), *Stirling* (EO\_TOU 4940), *frustrating* E2G\_FICTION 21515).

17092<sup>18</sup>: for them, why is no <Everlasting> Ice visible even in the  
17123: perwoman so that they can <bring> their giant ships filled  
17202: ' s pauses, the story is <telling> itself backwards. Perha  
17226: water, the story is too <exhausting> for her to lay out step  
17239: ponderous step and she is <dreaming> another moment far  
17323: rwards, but the story is <coming> backwards, it will even  
17365: as if waves of pain were <grinding> her into the spruce boug  
17379: bed. And finally one day <Greywing> begins to cry, even  
17400: lked motionlessly on this <lengthening> journey. Keskarrah  
17436: It ' s so long, " <Greywing> sobs, trying to burrow  
17439: long, " Greywing sobs, <trying> to burrow in. He pulls  
17553: " Gently Keskarrah rocks <Greywing> against himself " The

Fig. 11: Query excerpt from E2G\_FICTION\_EN

Retrieving information on gerunds in CroCo requires elaborate (semi-)automatic corpus queries which ideally rule out progressives, participles and words of other categories ending in *-ing* such as:

---

<sup>18</sup> The number indicated the Corpus ID, i.e. the position of the queried token in the corpus. Here the context of the query results has been set to several tokens, but not the whole sentence.

➤ **verbs**

bring, cling, fling, ring, sing, sling, spring, string, swing, wring  
(many of these verbs can also be used as nouns through conversion)

➤ **pronouns**

anything, everything, nothing, something

➤ **adjectives**

alarming, amazing, annoying, astonishing, booming, boring, charming, confusing, convincing, decreasing, depressing, disappointing, dying, embarrassing, exciting, existing, frightening, increasing, interesting, living, neighbouring, remaining, shocking, surprising, stunning, terrifying, thrilling, tiring, uninteresting, welcoming, worrying...

➤ **prepositions**

barring, concerning, considering, during, facing, following, including, notwithstanding, owing to, regarding, saving

➤ **nouns**

beginning, ceiling, darling, evening, fencing, feeling, heading, herring, king, morning, painting, plaything, pudding, shoestring, Thanksgiving, thing, ting, Viking, wedding, wing...

**surnames and geographical names:** Beijing, Ealing, Hawking, Irving, Peking, Reading, Stirling, Woking, Xiaoping...

**diminutives:** duckling, dumpling, gosling, nestling, seedling  
shilling, sibling, sterling, suckling, underling...

Theoretically it would be possible to exclude many irrelevant examples of other categories ending in *-ing* if they were all specified and listed in a query. The query however would look rather overloaded and unstructured<sup>19</sup> with the main goal of compensating for mistakes of the automatic analysis such as mistakes or omissions in the part-of-speech tagging. Writing such a query is also time-consuming and minor typing errors will result in syntax errors and slow down the process of querying. Many *-ing* forms are ambiguous. They could represent for example a noun, an adjective, a preposition or a gerund, depending on the context and the intended meaning.

One example of such an ambiguous case can be found in EO\_TOU 7284: *“flying wild geese”*. Here “to fly” can either mean “to fly in groups” - then “flying” would be an adjective - or “to shoot a bird in flight”, in which case it could be a gerund. More context is needed to understand the intended meaning. Human readers will make use of their world knowledge to interpret the sentence – something the computer programme is not capable of. Numerous compounds come up in CroCo, when

---

<sup>19</sup> e.g. [word!="a|an|very|more|most|t"] & pos!="VB.\*"] [word!="bring|sing|during|including" & word=".\*ing" & pos="V.\*"];

querying verbal *-ing* forms, some of them also inherently ambiguous (e.g. *developing countries* (E2G\_ESSAY: 7848), *operating companies* (E2G\_SHARE: 2372)). In many of these cases, human interpretation is required, a time-consuming process that is highly subjective and that might sometimes lead to disagreements, inconsistencies or misinterpretation of the results.

Excluding progressives via queries is an easier task. A query can rule out most progressives if it looks only for *-ing* forms that are not preceded by a form of “BE”<sup>20</sup>. The remaining progressives found in the query results are rather easy to spot and to strike out manually. They mainly occur if there are other elements, such as negations and adverbs, between the form of “BE” and the progressive or if several progressives follow each other. In questions, the subject is placed between “BE” and the progressive. The clitic „’s“ is not recognised by the programme as a form of „BE“ in queries. It should not be ruled out specifically in a query as it could be a genitive marker in some cases. “’m” is not recognised by the programme as a form of “BE” either.

Some progressives could be excepted if a preceding negation („not“ / „t“) was ruled out in a query. However, this would also exclude some gerunds after negations. Many adjectives could be ruled out by not allowing a preceding article or “more/most” and “very”. There are, nevertheless, some Ing-of gerunds after definite articles and we might lose some relevant examples in

---

<sup>20</sup> [pos!="VB.\*"] [word="\*.ing" & pos="V.\*"];

ambiguous cases or if the preceding words such as “BE” or “*most*” come at the end of a sentence and the next sentence starts with a gerund. In sum, a more complex query might generate finer grained results but the more cases we rule out specifically the higher the risk to lose relevant examples as well (“precision and recall” problem). Writing extremely complex queries anticipating all possible contexts and, at the same time, ruling out the highest possible number of irrelevant cases risks eventually to become more time-consuming than simply deleting irrelevant examples manually.

One of the biggest challenges however is the distinction between gerunds and participles, which is not easy either when done manually. A query that looks for *-ing* forms with verbal features that are not preceded by a form of “BE” in E2G\_ESSAY\_EN, for instance, leads to about 700 hits of which ca. 50% are irrelevant for our purposes because of their participle-status.

## **6.4.2 Investigating certain gerund patterns**

### **6.4.2.1 Ing-of gerunds and prepositions + gerunds**

Because of the complex aspects mentioned above that need to be considered in query design for gerunds, this study concentrates on investigating two typical patterns in which gerunds occur. It first looks at Ing-of gerunds (e.g. “*the signing of a protocol*”) in all registers of CroCo as they are relatively easy to query. Then the frequencies of the most important equivalent structures in the German translation corpus are analysed. Both the German options for translating English gerunds and the German structures that have been translated as gerunds in the G2E translation direction are analysed. The second pattern which this study focuses on are gerunds after prepositions, a topic about which De Smet (2008) already did some corpus-based research on the basis of the Helsinki Corpus where he focused on diachronic change in gerund constructions. Gerunds after prepositions in this study will also be examined in all corpus registers of CroCo as well as in the German translations from the sentence-aligned registers. Additionally, it will be shown which types of German structures have been translated by the use of gerunds in the G2E translation subcorpora. The query syntax and the query results of potential Ing-of gerunds and of potential preposition + gerund patterns in CroCo as well as the contexts of manually disambiguated and identified gerunds within these results and their sentence aligned

German equivalents have been saved and are available upon request.

**- ing + of**

<b>EO</b>			
	Query hits	after manual disambiguation	% (of relevant hits)
ESSAY	5	4	8
FICTION	13	8	16
SHARE	8	6	12
INSTR	8	6	12
POPSCI	18	13	25
SPEECH	8	6	12
TOU	7	1	2
WEB	11	7	13
$\Sigma$	78	51	100
<b>ETRANS</b>			
	Query hits	after manual disambiguation	% (of relevant hits)
ESSAY <sup>21</sup>	31	23	22
FICTION	8	2	2
SHARE	23	16	16
INSTR	35	26	25
POPSCI	21	17	17
SPEECH	16	9	9

<sup>21</sup> It should be noted that almost all sentences and their translation appear twice in the queries for G2E\_ESSAY, apparently a corpus error.



TOU	8	3	3
WEB	9	6	6
$\Sigma$	<i>151</i>	<i>102</i>	<i>100</i>

Fig. 12: Query results Ing-of gerunds in CroCo

## Preposition + -ing

<b>EO</b>	Query hits	after manual disambiguation	% (of relevant hits)
ESSAY	154	131	21
FICTION	28	20	3
SHARE	120	109	17
INSTR	91	85	14
POPSCI	80	72	11
SPEECH	122	110	18
TOU	35	15	2
WEB	95	86	14
$\Sigma$	725	628	100
<b>ETRANS</b>			
	Query hits	after manual disambiguation	% (of relevant hits)
ESSAY	195	147	22
FICTION	36	33	5
SHARE	148	124	18
INSTR	113	101	15
POPSCI	92	77	12
SPEECH	116	109	16
TOU	46	29	4
WEB	60	53	8
$\Sigma$	806	673	100

Fig. 13: Query results preposition + gerund in CroCo

### 6.4.2.2 Manual disambiguation and technical problems

As mentioned in Section 2.1, some gerunds of frequently used verbs seem to have established themselves as true nouns or seem to be on the verge of becoming nouns. Cases that were listed in a dictionary as nouns therefore were excluded later during manual disambiguation. We have mainly followed the category classification of the Longman Dictionary of Contemporary English. Nevertheless, the distinction between gerunds and nouns sometimes seems a bit arbitrary. There is no real syntactic difference between *the making of history / the breeding of animals* (listed as nouns in the dictionary) or *the sharing of profits* (not listed as a noun although *profit sharing / code sharing* are identified as a nouns). Words such as *understanding* are listed but not *preunderstanding* (because the word is probably too rare to have entered the dictionary). Examples of nouns with an additional prefix were sorted out manually in this study similarly to other nouns. Some of these examples are ambiguous and require careful reading and interpretation e.g. *funding* (money that is provided for a particular purpose → noun, or the process of giving money → gerund).

When looking for prepositions + *-ing*, several examples of conjunction + *-ing* were among the results and had to be sorted out manually, e.g.:

(154) Do not attempt to service this product yourself <as opening> or removing covers may expose you to dangerous (ETRANS\_INSTR 30096)

We can assume that not all relevant cases were found as the following example seems to prove:

(155) agreement or division of opinion ... . \_\_UNDEF\_\_  
<Because misunderstanding> and lack of understanding  
are only degenerate species (ETRANS\_POPSCI 2559)

Here in (155) a conjunction + *-ing* was found through querying but not the following preposition + *-ing* (although this example was sorted out anyway due to the noun status of *misunderstanding* and *understanding*). Probably in this case, the tagger – relying on statistical probabilities and POS-information from dictionaries – did not recognise the word *misunderstanding* and tagged it as a verb, whereas it recognized *understanding* as a noun.

Several words were tagged misleadingly as prepositions, e.g.:

- (156) *means* [wanting] (E2G\_ESSAY\_EN 12774),  
(157) *under* [lying principle] (G2E\_ESSAY\_EN 54765),  
(158) *debt* [servicing] (G2E\_SHARE\_EN 17013),  
(159) *Mozart* [whistling] (ETRANS\_TOU 38338),  
(160) *Boat* [Racing] (ETRANS\_WEB 19362)

The word “*worth*” was also tagged as a preposition in many cases (as in: to be worth doing). There is some debate in the literature on the category status of *worth*. Maling (1983) or the Longman dictionary for instance analyse *worth* as a preposition rather than an adjective. Huddleston & Pullum (2002) analyse it as an adjective. We have decided to sort these examples out from the query results. Cases where the distinction between particles and preposition in phrasal verbs (such as in: *go on doing*) was not entirely clear were not excluded from further investigation.

During the course of this study, it became clear that only three aligned registers of CroCo can be examined with CQP so far (ESSAY, FICTION, SHARE). That means that the translation of gerunds can only be analysed in three of eight registers. Another drawback is the strong similarity between the registers ESSAY and SHARE with regard to their business / politics vocabulary so that the query results are very alike. E2G\_SHARE\_EN has been aligned but apparently there were some problems as the query results show that almost no sentence in this subcorpus is connected to its translation. Almost all sentences and their translation appear twice in the register G2E\_ESSAY, a corpus error. Moreover, in the third register FICTION the distribution of gerunds in this particular register is generally low.

### **6.4.2.3 Query results and analysis**

Despite of problems mentioned above the analysis of the query results has led to several interesting findings (cf. Fig. 14-17).

### German equivalents of Ing-of gerunds (E2G)

	ESSAY		FICTION		SHARE	
	hits	%	hits	%	hits	%
infinitives	0	0	0	0		
other verbs	0	0	1	12.5	alignment	
<b>(<math>\Sigma</math> verbs)</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>12.5</b>	errors	
noun ending in -ung	2	50	2	25		
nominalised inf.	0	0	2	25		
other nouns	1	25	1	12.5		
<b>(<math>\Sigma</math> nouns)</b>	<b>3</b>	<b>75</b>	<b>5</b>	<b>62.5</b>		
adjective/adverb	1	25	0	0		
zero equivalent	0	0	0	0		
gerunds corresponding to other construction <sup>22</sup> or different vocabulary	0	0	2	25		
fragment too short, wrong alignment or completely different translation of sentence	0	0	0	0		
$\Sigma$	4	100	8	100	6	100

Fig. 14: German equivalents of Ing-of gerunds in CroCo (E2G)

<sup>22</sup> e.g. passive voice, verb – noun combinations / "Funktionsverbgefüge"

### German equivalents of Ing-of gerunds (G2E)

	ESSAY		FICTION		SHARE	
	hits	%	hits	%	hits	%
infinitives	0	0	0	0	0	0
other verbs	0	0	1	50	2	12
<b>(<math>\Sigma</math> verbs)</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>50</b>	<b>2</b>	<b>12</b>
noun ending in -ung	14	61	0	0	11	69
nominalised inf.	3	13	0	0	0	0
other nouns	3	13	1	50	3	19
<b>(<math>\Sigma</math> nouns)</b>	<b>20</b>	<b>87</b>	<b>1</b>	<b>50</b>	<b>14</b>	<b>88</b>
adjective/adverb	0	0	0	0	0	0
zero equivalent	0	0	0	0	0	0
gerunds corresponding to other construction or different vocabulary	0	0	0	0	0	0
fragment too short, wrong alignment or completely different translation of sentence	3	13	0	0	0	0
$\Sigma$	23	100	2	100	16	100

Fig. 15: German equivalents of Ing-of gerunds in CroCo (G2E)



### German equivalents of preposition + gerund (E2G)

	ESSAY		FICTION		SHARE	
	hits	%	hits	%	hits	%
infinitives	12	9	8	40		
other verbs	26	20	8	40	alignment	
<b>(<math>\Sigma</math> verbs)</b>	<b>38</b>	<b>29</b>	<b>16</b>	<b>80</b>	errors	
noun ending in -ung	46	36	0	0		
nominalised inf.	3	2	0	0		
other nouns	16	12	2	10		
<b>(<math>\Sigma</math> nouns)</b>	<b>65</b>	<b>50</b>	<b>2</b>	<b>10</b>		
adjective/adverb	0	0	0	0		
zero equivalent	10	8	1	5		
gerunds corresponding to other construction or different vocabulary	16	12	1	5		
fragment too short, wrong alignment or completely different translation of sentence	1	1	0	0		
$\Sigma$	130	100	20	100	109	100

Fig. 16: German equivalents of prepositions + gerunds in CroCo (E2G)

### German equivalents of preposition + gerund (G2E)

	ESSAY		FICTION		SHARE	
	hits	%	hits	%	hits	%
infinitives	29	20	14	43	19	17
other verbs	23	16	5	15	11	9
<b>(<math>\Sigma</math> verbs)</b>	<b>52</b>	<b>36</b>	<b>19</b>	<b>58</b>	<b>30</b>	<b>26</b>
noun ending in -ung	34	23	3	9	30	26
nominalised inf.	0	0	2	6	2	2
other nouns	10	7	1	3	11	9
<b>(<math>\Sigma</math> nouns)</b>	<b>44</b>	<b>30</b>	<b>6</b>	<b>18</b>	<b>43</b>	<b>37</b>
adjective/adverb	2	1	1	3	5	4
zero equivalent	30	20	4	12	25	23
gerunds corresponding to other construction or different vocabulary	11	7	3	9	6	5
fragment too short, wrong alignment or completely different translation of sentence	8	6	0	0	6	5
$\Sigma$	147	100	33	100	115	100

Fig. 17: German equivalents of prepositions + gerunds in CroCo (G2E)

The distribution of Ing-of gerunds is relatively low (E2G: 51, G2E: 102) compared to prepositions + gerunds (E2G: 628, G2E: 673). The results of the investigation of prepositions + gerunds are therefore more significant and representative than the results of Ing-of gerunds.

Most Ing-of gerunds were found in the following registers:

E2G: POPSCI (13), FICTION (8), WEB (7)

G2E: INSTR (26), ESSAY (23), POPSCI (17)

Ing-of gerunds had the lowest frequencies in the following registers:

E2G: TOU (1), ESSAY, (4) SHARE/INSTR/SPEECH (6)

G2E: FICTION (2), TOU (3), WEB (6)

Most gerunds after prepositions were found in:

E2G: ESSAY (131), SPEECH (110), SHARE (109)

G2E: ESSAY (147), SHARE (124), SPEECH (109)

These patterns had the lowest frequencies in:

E2G: TOU (15), FICTION (20), POPSCI (72)

G2E: TOU (29), FICTION (33), WEB (53)

Fig. 18: Highest and lowest gerund frequencies in CroCo registers

If we focus on the more representative results of prepositions + gerunds we come to the conclusion that these gerunds patterns are more frequent in original texts and translations of ESSAY, SPEECH and SHARE than in other registers and less frequent in TOU and FICTION than in other registers. However, as mentioned above, many passages appear twice in G2E\_ESSAY.

Surprisingly, most registers in G2E are characterised by a considerably higher number of both examined gerund patterns than in E2G. One could have expected to find more of these gerund patterns in English originals than in English translations. Translated texts are assumed to be influenced by source language interference. In many cases, translators tend to imitate grammatical patterns of original texts where this is possible („shining-through“ in E-G translations, cf. Teich, 2003). Texts translated from German to English could be expected that NPs and VPs are often translated by NPs and VPs instead of gerunds when selecting from several translation possibilities. Nevertheless, the higher number of gerunds in G2E seems to indicate that translators actually prefer gerunds when they can chose from grammatically different options. Hopkinson (2007) made a similar observation for a Czech-English translation corpus. English translations from Czech texts tend to have as many gerunds or even more than English originals.

There might be a language-inherent explanation for this phenomenon. It has often been said that the English language has a stronger verbal orientation compared to languages like German

(Hansen-Schirra et. al, forthcoming, p.76) and using many gerunds, derived from verbs, could be a way of avoiding noun-heavy syntax.

It is rather difficult to say from the analysis made about certain gerund patterns whether the original texts are more verbal than their translations or whether the English texts are more verbal than the German texts. English gerunds themselves are on the edge between nouns and verbs and they correspond frequently to nouns or verbs in the German texts in CroCo.

The higher use of gerunds in translations may also be due to common translational processes and specific properties of translated texts that distinguish these texts from non-translated ones. Baker identified four main translation universals: explicitation, simplification, normalisation und levelling out (Baker, 1996, cf. Steiner, 2004).

According to Baker translators tend to make information of the source text more explicit in the target text. They consciously or subconsciously also use a simpler style in translations than the corresponding original texts. Normalisation means that culture specific aspects are taken into account, fragments are completed and untypical or erroneous structures replaced by unmarked and correct ones. Levelling out refers to the observation that the lexical density and average text length of translations is different from that of non-translated texts.

The overuse of gerunds in translations seems to prove a tendency towards normalisation and generalisation, as translators

unconsciously seek regularity in language systems and tend to favour processes which they see as being 'regular', i.e. the transformation of English verbs into gerunds by the addition of the *-ing* suffix.

For example, there is no exact equivalent structure for gerunds in Czech. The most similar forms are verbal nouns. According to Hopkinson, translators use *-ing* forms in many cases where English would avoid the gerund and instead use a common nominal form. In translations, Czechs sometimes overuse patterns like „*the demolishing of*“ instead of „*the demolition of*“ which can be seen as an attempt to imitate Czech verbal nouns. The Czech verb “*zbourat*” (to demolish) for instance can be transformed into the verbal noun “*zbourání*”.

In German translations, something similar can be observed. In (161), the translator decided to translate the German noun “*DM-Schwäche*” with a gerund implying a process: “*the weakening of the D-Mark*” although the NP “*the weakness of the D-Mark*” or a shift towards an adjective (“*the weak D-Mark*”) would have been equally good translations with regard to content and style. The phrase “*ein kleiner Beitrag zur Aufklärung*” in (162) became “*a small contribution towards educating the public*”, but “*a small contribution to education*” would be as good.

(161) G2E\_SHARE 10845: The <weakening of> the D-Mark  
compounded the price increase; ...

-->g2e\_share\_ge: Die DM-Schwäche machte den  
Preisanstieg noch gravierender, ...

(162) G2E\_FICTION 1440: I calmed down and tried to  
convince myself that an act like that, if I committed it,  
would be nothing but an act of good, a small contribution  
<toward educating> the public, toward democracy, toward  
justice.

-->g2e\_fiction\_ge: Ich wurde ruhiger und versuchte mir  
einzureden, dass eine solche Tat, wenn ich sie beginge,  
nichts weiter als eine gute Tat sei, ein kleiner Beitrag zur  
Aufklärung, zur Demokratie, zur Gerechtigkeit.

In other cases, gerunds were preferred in the translation  
although infinitives would also have been possible (163).

(163) 66195: This strategy makes science nothing less than  
the key <to halting> the depopulation of east Germany.

-->g2e\_essay\_ge: Die Wissenschaft wird damit zugleich  
zum Schlüssel, die Entvölkerung des Ostens aufzuhalten.

Very often, translators preferred structures that make gerunds obligatory (164-167). Particularly in ESSAY and SHARE, various German verbs have been translated by *succeeded in + gerund* or *be successful in + gerund* although alternatives with an infinitive were also possible such as *we managed to / this has enables us to / we have been able to / this made it possible to + infinitive*. This seems to prove a tendency towards simplification in translations, which are assumed to have fewer linguistic realisation devices.

(164) 179: As a result, we have succeeded <in eliminating> all group losses [...]

-->g2e\_share\_ge: So ist es uns gelungen, alle Verlustquellen des Konzerns zu schließen [...]

(165) 13190: [...] thanks – not least – to the total commitment of the Lufthansa staff and their willing acceptance of restraints on pay, we have succeeded <in returning> an operating profit of Euro 718m for the year 2002.

-->g2e\_share\_ge: [...] volles Engagement und ein Gehaltsverzicht der Lufthanseaten haben es ermöglicht, im Jahr 2002 ein operatives Ergebnis von 718 Mio. Euro zu erwirtschaften.



- (166) 31584: As a result of our cost management we succeeded <in limiting> the growth in administrative expense.  
 -->g2e\_share\_ge: Durch unser Kostenmanagement wurde die Zunahme des Verwaltungsaufwandes begrenzt.
- (167) 19389: It has succeeded <in combining> economic prosperity with social justice [...] [...]  
 -->g2e\_essay\_ge: Sie hat wirtschaftlichen Wohlstand und soziale Gerechtigkeit miteinander verbunden [...]

Another noteworthy observation is the considerable number of cases where gerunds were inserted by translators and had no explicit lexical or grammatical equivalent in the German original text. In translations from English to German, gerunds were omitted from time to time making the translation more implicit. However, the tendency to make the translation more explicit was more pronounced (zero equivalents: G2E\_ESSAY: 20% vs. E2G\_ESSAY only 8%; G2E\_FICTION: 12% vs. E2G\_FICTION only 5%, cf. [168-170]).

- (168) 21969: As well <as providing> direct help for his family,  
 [...]
 -->g2e\_share\_ge: Neben der unmittelbaren Hilfe für die  
 Hinterbliebenen [...]
- (169) 1434: [...] we will continue to do our share <in boosting>  
Jena spirit, [...]  
 -->g2e\_share\_ge: Und auch in Zukunft werden wir zum  
Jena-Spirit beitragen,
- (170) 5436: While there has been significant progress <in  
achieving> cyclical convergence, [...]  
 -->e2g\_essay\_ge: Trotz erheblicher Fortschritte bei der  
zyklischen Konvergenz [...]

In sum, the closest German equivalents of gerunds are infinitives, nominalised infinitives, *ung*-nouns and other deverbal nouns. In translations, these forms often become gerunds instead of nouns and infinitives. This might be due to the fact that particularly the use of infinitives is often blocked by prepositions in English. Nevertheless, another contributing factor is probably the common recommendation in translation training to use more gerunds in order to avoid the German noun-heavy style and unnecessary complexity in subclauses (e.g. Mautner, 2008). This recommendation is probably not based on statistical findings in

linguistic corpora but rather a general one. However, it shows that translation students are encouraged to replace other structures by gerunds. Remarkably, using many gerunds is even a common advice to native speakers of English who want to improve their written style. In an academic writing tip by an American university for example it is claimed that because English is a verbal language students need to replace nominal structures by verbal ones: “Whenever possible use strong verbs and replace nouns with verbs. Also use gerunds instead of nouns and prepositions.”<sup>23</sup> Apparently, English native speakers need to be encouraged to use gerunds more often.

In CroCo, generally more Ing-of gerunds and prepositions + gerunds were found in English translations than in English originals but a closer look on the subcorpora puts these numbers a bit into perspective. Firstly, the results show that the percentile distribution of prepositions + gerunds is very similar in E2G and G2E (e.g. ca. 21% of prepositions + gerunds were found in EO\_ESSAY and in ETRANS respectively, ca. 17% in EO\_SHARE and ETRANS\_SHARE and so on). The results of the investigation of Ing-of gerunds show slightly less similarity but these examples are less extensive.

It can be argued that the English originals were a bit shorter than the German originals with regard to the number of sentences

---

<sup>23</sup> <http://www.ndu.edu/ismo/docUploaded/Verb%20Suppression.pdf>  
(last accessed 8 August 2010)

and in some cases also with regard to the number of tokens. The English originals were also shorter than the English translations with regard to absolute numbers of sentences and tokens.<sup>24</sup>

	tokens <sup>25</sup>	sentences
EO_ESSAY	34,998	1,497
EO_FICTION	36,996	1,800
EO_SHARE	35,824	1,477
EO_INSTR	36,167	2,461
EO_POPSCI	35,148	1,422
EO_SPEECH	35,062	1,573
EO_TOU	35,907	1,489
EO_WEB	36,119	1,654
$\Sigma$	286,221	13,373
GO_ESSAY	35,668	1,903
GO_FICTION	36,778	2,155
GO_SHARE	35,235	1,805
GO_INSTR	36,880	2,601
GO_POPSCI	36,177	1,605
GO_SPEECH	35,337	2,002
GO_TOU	36,574	1,927

---

<sup>24</sup> It might have been slightly more accurate to compare the number of clauses instead of sentences but these data were not available for all registers due to a loss of that data in some files.

<sup>25</sup> Number of tokens including punctuation marks

GO_WEB	35,779	1,988
$\Sigma$	288,428	15,986
ETRANS_ESSAY	42,036	1,780
ETRANS_FICTION	40,037	2,107
ETRANS_SHARE	39,511	1,765
ETRANS_INSTR	39,663	2,553
ETRANS_POPSCI	37,878	1,566
ETRANS_SPEECH	39,766	1,937
ETRANS_TOU	43,677	1,794
ETRANS_WEB	39,657	1,882
$\Sigma$	32,2225	15,384

Fig. 19: Tokens and sentences per register

The results in Fig. 20 show that the examined gerund patterns occur roughly as often or slightly more often in EO as in ETRANS.

EO		
Ing-of gerunds / token	51/286,221	0.0002
Ing-of gerunds / sentence	51/13,373	0.0038
preposition + gerund / token	629/286,221	0.0022
preposition + gerund / sentence	629/13,373	0.0470

ETTRANS		
Ing-of gerunds / token	102/322,225	0.0003
Ing-of gerunds / sentence	102/15,384	0.0066
preposition + gerund / token	673/322,225	0.0021
preposition + gerund / sentence	673/15,384	0.0438

Fig. 20: Ratio of gerunds per tokens and sentences

The queries in CroCo reveal some interesting findings about German equivalents for gerunds in general. Section 6.3 presented a summary of previous corpus-corpus-based studies on gerunds. Among the few corpus-based studies that have investigated the translation of gerunds are those by Mehl, examining the frequencies of certain translations for gerunds in German.

Although Mehl examined all gerunds in a corpus, and not only certain types, the results in this study are finer-grained in many regards. As his corpus served MT-purposes, the corpus was rather small with about 100 examples of gerunds overall and he only examined the translation from English to German in one particular register. In his corpus, he found that all gerunds were either translated as verbs (65%), common nouns (17%) or nominalised infinitives (18%). He explained in great detail the reasons for choosing one variant over the other, such as ambiguity, style or the availability of nouns which are morphologically related to verbs in the gerund form.

The CroCo corpus is a larger and more representative corpus with different registers, translated and original English texts and their respective German equivalents. It provides more generalisable results and shows how often English gerunds correspond to German infinitives or other verbs, nouns ending in *-ung*, nominalised infinitives or other nouns, verb-noun combinations, adjectives or adverbs and zero equivalents etc.

Queries of both Ing-of gerunds and of gerunds after prepositions show that in ESSAY 50% or more of these gerunds have been translated as German nouns. Similarly, German nouns in ESSAY and SHARE were often translated with gerunds. The majority of these gerunds in English translations were German nouns ending in *-ung*, the rest were mostly other deverbal nouns. Nominalised infinitives and adjectives or adverbs corresponding to gerunds were rare in all examined registers.

In FICTION, on the other hand, gerunds correspond to German verbs in at least 50% of all cases. The explanation for these results is evident. ESSAY and SHARE are generally characterised by a noun-heavy syntax whereas fictional texts use more verbs.

#### 6.4.2.4 Particular characteristics of individual registers

This section highlights some register-specific characteristic patterns in the individual CroCo corpus registers. Depending on the text genre, gerunds often occur in the same patterns. Typical gerund triggers in ESSAY are: *be committed to...*, *to aim at...*, *look forward to...*, ... *is essential*.

Irrelevant examples that had to be sorted out manually in ESSAY also frequently fell into certain categories:

- compounds such as: *developing countries*, *annual meeting*, *managing director*
- ‘gerunds’ that have established themselves as nouns: *founding*, *training*, *polling*, *trading*
- adjectives: *existing*, *differing*, *continuing*

Similar findings can be observed in SPEECH and SHARE.

In INSTR, the preposition before gerunds is nearly always ‘by’, several compounds proved to be irrelevant examples like in ESSAY, SPEECH and SHARE such as: *weighing mode*, *measuring points*, etc. Formal language using many gerunds and restricted vocabulary to ensure unambiguity of the messages are typical for texts in INSTR, e.g. *by pressing*, *clicking*, *running*, *holding*, *using*, *starting*... Nevertheless, due to the mixed-category status of gerunds it is not entirely clear whether the frequent use of gerunds in English accounts for the degree of formality or a noun-heavy syntax.



Texts in FICTION and POPSCI have a much richer vocabulary and a high frequency of proper names, some of them wrongly tagged as verb (VVG), which makes them potential candidates for gerunds or participles in the query results (ETRANS\_POPSCI 6586: <Notting> Hill, ETRANS\_POPSCI 36386: <Withering>'s day). Several place names were also tagged as verbs in TOU (ETRANS\_TOU 15391 *Haring*, EO\_TOU 26079 *Reading*, ETRANS\_TOU 39427 *Grinzing*).

Texts in TOU were similar to ESSAY in one particular way due to the high numbers of ‘gerunds’ that have established themselves as nouns, particularly free time activities such as *camping, skiing, hiking, cycling, sail-flying etc.*

## **7 Conclusions and suggestions for further research**

This study has led to some interesting findings about translation strategies for a grammatical pattern that has no exact equivalent in another language. It has been shown that the gerund is a peculiarity of the English language which has developed in a highly complex process under the influence of various language-internal and external factors. A semi-automatic query mechanism for certain gerund patterns in a language corpus has been developed after the consideration and discussion of technical problems.

In sum, all corpus-based studies on gerunds faced the same challenge so far: The lack of commonly accepted definitions among linguists, the mixed-category status of gerunds between nouns and verbs and the fact that gerunds occur in various patterns and functions make the identification of gerunds in a corpus rather difficult compared to other categories. In previous corpus-based studies on gerunds usually all *-ing* forms were queried and then disambiguated manually. This procedure is only feasible if the corpus is not particularly large or if only certain gerund patterns are taken into account. Therefore it has not been possible to query all gerunds in a corpus as big as the CroCo corpus.

In this thesis, register differences have been shown and evidence for translation specific properties could be found. It could also be shown that the analysed gerund patterns occurred at

least as often in English translations as in original texts. Among other things, it might be interesting to further investigate the influence of the native language of translators for the German-English language pair on the use of gerunds, when they are translating into their mother tongue or into a foreign language.

In a future study, more gerund patterns might be queried, particularly with regard to their grammatical function as subject or object for example. This study focussed only on two gerund patterns to test whether a linguistic corpus is suitable to examine gerunds and their translation quantitatively. Investigating gerunds of certain grammatical functions would be feasible in the CroCo corpus. However, it would require more complicated queries looking for certain sequences of characters with a certain word class and grammatical function. This study could serve as a basis for future research on that topic.

## References

Abney, S. (1987). *The English Noun Phrase in its Sentential Aspect*, PhD thesis, MIT. MIT Working Papers in Linguistics.

Algeo, J. (2010). *The Origins and Development of the English language*. (6th ed.). Wadsworth: Cengage Learning.

Andor, J. (2004), *The Master and his Performance: An Interview with Noam Chomsky*, *Intercultural Pragmatics*, 1-1, Walter de Gruyter, pp. 93-111.

Atkins, S., J. Clear & N. Ostler (1992). *Corpus Design Criteria*. *Literary & Linguistic Computing* 7:1, pp. 1-16.

Baker, M. (1985). *Syntactic affixation and English gerunds*. In *Proceedings of the West Coast Conference on Linguistics* 4, 1-11. Stanford, Calif.: CSLI Publications.

Baker, M. (1996) 'Corpus-based Translation Studies. The Challenges that Lie Ahead', in Harold Somers (ed) *Terminology, LSP and Translation*, Amsterdam: John Benjamins, pp. 175-86.

Biber, D. (1993). *Representativeness in Corpus Design*. *Literary & Linguistic Computing* 8:4, pp. 243-257.

Biber, D., S. Conrad, & R. Reppen (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge UP.

Broderick, J.P. (2000). *Toward a 'monosyntactic' (and monosemic?) analysis of -ing forms in English*. Address delivered at the Sixty-third Meeting of the Southeastern Conference on Linguistics (SECOL LXIII), Birmingham, AL, November 11, 2000. <http://courses.lib.odu.edu/engl/jpbroder/2000d%5Bingforms%5D.doc> (last accessed 8 August 2010)

Catford, J.C. (1965). *A Linguistic Theory of Translation*. Oxford University Press, Oxford.

Cho, J.Y. (1985). On Diachronic Change of ‘- ing Form’ in OE and ME. <http://mlsk.or.kr/main/data2/mlsk2008090320473961.PDF> (last accessed 8 August, 2010)

Chomsky, N. (1962), paper given at Third Texas Conference on Problems of Linguistic Analysis in English, 1958, p. 159. Austin: University of Texas.

Comrie, B. (1976). The syntax of action nominals. A cross-linguistic study. *Lingua* 40, pp. 177–201.

Comrie, B. & S.A. Thompson. (1985). Lexical nominalization. Timothy Shopen (ed.) *Language Typology and Syntactic Description*, Vol. 3. Cambridge: Cambridge University Press, pp. 349-98.

Croft, W. (2000). *Explaining language change. An evolutionary approach*. London: Longman.

Curme, O. (1980). *A Grammar of the English Language, Volume II: Syntax*. Essex, CT, Verbatim.

Cyrus, L. (2006). Building a resource for studying translation shifts. In *Proceedings of LREC 2006 – Second International Conference on Language Resources and Evaluation*, Genoa, Italy.

Evert, S. (2005). *The CQP Query Language Tutorial*. University of Stuttgart.  
[http://www.ims.unistuttgart.de/projekte/CorpusWorkbench/CQP\\_Tutorial/cqp-tutorial.pdf](http://www.ims.unistuttgart.de/projekte/CorpusWorkbench/CQP_Tutorial/cqp-tutorial.pdf) (accessed 8 August 2010)

Expósito González, M. (1996). *La estructura del sintagma nominal en el inglés de la cancillería: 1400-1450*. Barcelona: Kadle Books.

Fanego, T. (1996). The gerund in Early Modern English. Evidence from the Helsinki Corpus. *Folia Linguistica Historica* 17, pp. 97–152.

Fanego, T. (2004). On reanalysis and actualization in syntactic change: the rise and development of English verbal gerunds. *Diachronica* 21/1, pp. 5-55.

Fowler, H.W. (1908) *The King's English*. 2nd ed. Chapter II. Syntax: Participle and gerund, Oxford: Clarendon press, 1908. New York: bartleby.com, 1999.  
<http://www.bartleby.com/116/212.html> (last accessed 8 August 2010)

Halliday, M.A.K., & R. Hasan (1989). *Language, Context and Text: Aspects of language in a social-semiotic perspective*. Geelong, Victoria: Deakin University Press. (republished by Oxford University Press 1989).

Hansen-Schirra, S., S. Neumann & M. Vela (2006). Multi-dimensional Annotation and Alignment in an English-German Translation Corpus. *Proceedings des Workshops Multidimensional Markup in Natural Language Processing (NLPXML-2006)* Trento, Italy, 4 April 2006, pp. 35-42.

Hansen-Schirra, S., S. Neumann & E. Steiner (forthcoming). *Cross-linguistic Corpora for the Study of Translations. Insights from the Language Pair English-German*. Berlin, New York: de Gruyter.

Houston, A. (1989). The English Gerund: Syntactic Change and Discourse Function. In: Fasold, R.W. / Schiffrin, D. (eds) *Language Change and Variation*. Amsterdam: Benjamins, pp. 173-196.

Huddleston, R. D. & G. Pullum (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press.

Hudson, R. (2003). Gerunds without phrase structure. *Natural Language and Linguistic Theory* 21, pp. 579-615.

Jasanoff, J.H. (2006). The origin of the Latin gerund and gerundive: a new proposal. In Harvey Goldblatt and Nancy Kollman (eds), *A Festschrift in Honor of Michael Flier* [= *Harvard Ukrainian Studies* 28].

Jespersen, O. (1933). *Essentials of English Grammar*. London: Allen & Unwin.

Kadeeri, H.K. (2008). Gerunds in Shakespearean English. *Kagoshima studies in English literature* Vol.17, pp. 1 -16

Kennedy, G.D. (1998). *An Introduction to Corpus Linguistics*. London: Longman.

König, E. & V. Gast (2009). *Understanding English-German Contrasts*. 2<sup>nd</sup> rev. ed. *Grundlagen der Anglistik und Amerikanistik*, no 29. Berlin: Erich Schmidt Verlag.

Königs, K. (2004). *Übersetzen Englisch-Deutsch. Ein systemischer Ansatz*. München: Oldenbourg, 2<sup>nd</sup> rev. ed.

Lees, R. (1960). *The Grammar of English Nominalizations*. The Hague: Mouton.

Longman / Pearson (ed.) (2009). *Longman Dictionary of Contemporary English*.

Lyne, S. (2006): The Form of the Pronoun Preceding the Verbal Gerund: Possessive or Objective?, *ICAME Journal* 30, pp. 37-53.

Lyne, S. (2007). 'There is no point in our wasting any more time' – or 'us wasting...'? Variation between genitive and non-genitive subjects of verbal gerunds in Present-Day English. Paper presented at the Oxford Postgraduate Linguistics Conference (LingO), 21/10/2007.

Lyons, J. (1977). *Semantics*. Vol. I,II. Cambridge Univ. Press.  
German Translation: *Semantik*. München: Beck, Vol. I: 1980, Vol. II: 1983.

Maling, J. (1983). 'Transitive Adjectives: A Case of Categorical Reanalysis'. In F. Heny and B. Richards (eds.), *Linguistic Categories: Auxiliaries and Related Puzzles*, Vol. 1, Dordrecht: Reidel, pp. 253-289.

[http://people.brandeis.edu/~maling/Maling1983\\_adjectives.pdf](http://people.brandeis.edu/~maling/Maling1983_adjectives.pdf)  
(last accessed 8 August 2010)

Manolessou, I. (2005). From Participles to Gerunds. In *Advances in Greek Generative Syntax*, Stavrou, Melita and Arhonto Terzi (eds.), pp. 241–283.

Mautner, G. (2008). *Englische Grammatik für die Wirtschaftskommunikation*. 2<sup>nd</sup> rev. ed., Vienna: Linde.

<http://www.wu.ac.at/ebc/downloads/grammar.pdf> (last accessed 8 August 2010)

McEnery, T. & A. Wilson (2001). *Corpus Linguistics – An Introduction* (2nd ed.). Edinburgh: Edinburgh UP.

Mehl, S. (1995). Interaction between Syntax and Semantics: The Case of Gerund Translation, in K. de Smedt (ed.), *Proceedings of the 5th European Workshop on NLG*, Leiden

Mehl, S. (1996). Systematic Alternatives in Lexicalization: The Case of Gerund Translation. *Machine Translation* 11, pp. 185-216.

Mehl, S. & M. Volk (1999). Aspects of the Translation of English Subordinate Clauses into German. In: *Problems and Potential of English-to-German MT systems*. Proceedings of Workshop at the 8th International Conference on Theoretical and Methodological Issues in Machine Translation, Chester.

Miller, G. (2002). *Nonfinite Structures in Theory and Change*. Oxford: Oxford University Press. (Chapter10: English Gerundials, Chapter 11. The History of English Gerundials).



Neumann S. & S. Hansen-Schirra (2005). The CroCo Project. Cross-linguistic Corpora for the Investigation of Explicitation in Translations. Proceedings of the Corpus Linguistics Conference Series Vol. 1 (1).

Neumann, S. (2008). Contrastive Register Variation. A Quantitative Approach to the Comparison of English and German. Habilitationsschrift., Saarland University.

Núñez-Pertejo, P. (1996). On the Origin and History of the English Prepositional Type A-Hunting: A Corpus-Based Study. *Revista Alicantina de Estudios Ingleses* 9, pp. 105-18.

Núñez-Pertejo, P. (2004). The Progressive in the History of English with Special Reference to the Early Modern English Period: A Corpus-Based Study. Munich: LINCOS Studies in English Linguistics 06.

Oxford University Press (ed.) (1971). The Compact Edition of the Oxford English Dictionary.

Pearson, J. (1998). *Terms in Context*. Amsterdam: John Benjamins Publishing Company.

Poutsma, H. (1929). *A Grammar of Late Modern English*, Groningen: Noordhoff.

Quirk, R. et al (1985). *A Comprehensive Grammar of the English Language*. Longman.

Reuland E. (1983). Governing -ing, *Linguistic Inquiry* 14.1.

Rosch, E. (1973). On the Internal Structure of Perceptual and Semantic Categories. In: Moore, Timothy E. (ed.): *Cognitive Development and the Acquisition of Language*. New York, pp. 111-144.

Ross, J.R. (1973). Nouniness. In: Fujimura, O. (ed.). *Three Dimensions of Linguistic Theory*. Tokyo: Tokyo Institute for Advanced Studies of Language.

Sampson G. (1995). *English for the Computer*. Oxford, Oxford University Press.

de Smet, H. (2008) Functional motivations in the development of nominal and verbal gerunds in Middle and Early Modern English. *English Language and Linguistics* 12, pp. 55-102.

Steiner, E. (2004). *Translated Texts. Properties, Variants, Evaluations*. Frankfurt/M.: Peter Lang.

Smitherberg, E. (2005). *The Progressive in 19th-century English: A Process of Integration*. *Language and Computers: Studies in Practical Linguistics* 54. Amsterdam and New York: Rodopi.

Teich E., (2003). *Cross-linguistic Variation in System and Text*. Mouton de Gruyter, Berlin and New York.

Tikkanen, B. (1987). *The Sanskrit Gerund: A Synchronic, Diachronic and Typological Analysis*. *Studia Orientalia* (Helsinki) 62.