Chapter 5

An analysis of translational complexity in two text types

Martha Thunes
University of Bergen

This article presents an empirical study where translational complexity is related to a notion of computability. Samples of English-Norwegian parallel texts have been analysed in order to estimate to what extent the given translations could have been produced automatically, assuming a rule-based approach to machine translation. The study compares two text types, fiction and law text, in order to see how these differ with respect to the question of automatisation. A central assumption behind the empirical method is that a specific translation of a given source expression can be predicted, or computed, provided that the linguistically encoded information in the original, together with information about source and target languages, and about their interrelations, provides the information needed to produce that specific target expression. The results of the investigation indicate that automatic translation tools may be helpful in the case of the law texts, and the study concurs with the view that the usefulness of such tools is limited with respect to fiction. Finally, an extension of the analysis method is proposed in order to make it relevant as a diagnostic tool for the feasibility of automatic translation in relation to specific text types.

1 Introduction

The present contribution is based on an empirical study of translational correspondences identified in selected English-Norwegian parallel texts. Two main research questions will be discussed, and the first one is about automatisation: i.e., to what extent is it possible to automatise, or compute, the actual translation relation found in the investigated parallel texts? The study attempts to answer this by analysing the parallel texts into pairs of translationally corresponding

units, primarily at clause level, and measuring the degree of translational complexity in each such correspondence. In the investigated material, the target texts have been produced by human translators.

The second research question deals with text type. The data include two text types, fiction and law text, and these have been compared in order to find out if there is, in the empirical material, a difference in the degree of translational complexity between the two text types. In relation to this second question, an important factor is the difference in the degree of restrictedness between fiction and law text.

1.1 Key concepts

The applied notion of translational complexity is defined in terms of the types and amount of information needed when a specific translation is produced from a given source expression. Since this conception of translational complexity is related to linguistic information, the present study is seen as relevant to linguistic approaches to machine translation (MT), commonly known as rule-based MT (RBMT).[1]

In this study, 'automatisation' is understood simply as the generation of translations with no human intervention, but the investigation is not related to any particular translation algorithm or system architecture. Rather, the intention is to discuss automatisation with reference to information about languages by relating it to an assumption concerning predictability in the translational relation. I.e., we assume that there is a translational relation between the inventories of simple and complex linguistic signs in two languages which is predictable, and then also *computable*, from information about source and target language systems, and about how the languages correspond.

This means that a computable translation is *linguistically predictable*, i.e. predictable as one of possibly several alternative translations, and the basis for predicting it is the linguistic information coded in the source text, together with given, general information about the two languages and their interrelations.[2] It also means that *non-computable* translations cannot be predicted merely from these types of linguistic information, because non-computable translation tasks

---

[1]*Rule-based MT* is the classic approach to machine translation, where the translation procedure relies on information about source and target language and their interrelations, and this is in contrast to *statistical MT* (SMT), or modern machine translation, where translations are computed on the basis of statistical information about existing correspondences in large bodies of parallel texts. See Jurafsky & Martin (2009: 898).

[2]Cf. Dyvik (1998: 52) on the notion of linguistically predictable translation.

require access to additional information sources, such as various kinds of general or task-specific extra-linguistic information, or task-specific linguistic information from the context surrounding the source expression.

In order to answer the research questions given in section 1, a measurement of *translational complexity* is applied to the analysed texts. For this purpose, pairs of translationally corresponding linguistic units, primarily finite clauses, are identified as individual *translation tasks*, and 'translational complexity' is defined in the following way: in a given translation task, the degree of translational complexity is a factor determined by the types and amount of information needed to solve the task, as well as by the accessibility of these information sources, and the effort required when they are processed. The analysis to be presented is carried out within a strictly product-oriented approach; aspects related to translation methods, or to the cognitive processes behind translation, will not be considered.

In the present approach, a scale of translational complexity is assumed, and, for analytical purposes, four main types of translational correspondences are identified on this scale. The four correspondence types are organised in a hierarchy, reflecting an increase in the degree of translational complexity. Moreover, the issue of computability is closely related to the categorisation of translational correspondences. That is, a dividing line between computable and non-computable translation tasks can be drawn on a certain point across the scale of translational complexity.

1.2 Outline

This article is organised in the following way: §2 presents related approaches to the classification of translational correspondences, and discusses points of contact between the present work and, respectively, translation studies and machine translation. In §3 the correspondence type hierarchy is explained and illustrated, and some of its underlying assumptions are commented on. §4 describes how the classification model has been applied to the investigated data; it presents the analysed parallel texts, as well as certain text-typological aspects, and gives the results of the empirical analysis. §5 discusses the results in relation to the research questions given in §1, and comments further on the relevance of this study for automatic translation.

2 Related works

The type hierarchy to be presented in §3 is a fairly general classification model for translational correspondences, and its basic principles were originally defined by Helge Dyvik of the University of Bergen.[3] A further development of his model is previously published in Thunes (1998), and the approach applied in the present study is described in more detail in Thunes (2011). The model has also been adopted by several other researchers within contrastive language studies. For the purpose of analysing word-order differences between English and Norwegian, Hasselgård (1996) employs a slightly modified version of the correspondence type hierarchy as defined by Dyvik (1993), and her approach is further developed in an English-Norwegian study of thematic structure (Hasselgård 1998). Elgemark (2017) has adapted the analytical approach of Hasselgård (1998) to a contrastive study of clause-final constituents in English-Swedish. Modified versions of the correspondence type hierarchy as presented in Thunes (1998) are used by Tucunduva (2007); Silva (2008), and Azevedo (2012), all of which are studies where the model is applied for the purpose of analysing and describing translational correspondences in English-Portuguese parallel texts.

Other related approaches are found in the works of, respectively Merkel (1999); Cyrus (2006), and Macken (2010). These contributions are rooted in computational linguistics, in addition to being of relevance to contrastive language studies and translation research. Merkel (1999) presents a model for the description of structural and semantic correspondences in Swedish-English parallel texts. Cyrus (2006) develops a framework for manual annotation of translationally interrelated predicate-argument structures in a German-English parallel corpus. With reference to Dutch-English, Macken (2010) presents research on automatic alignment of translational correspondences below sentence level.

The type hierarchy model of the present study may be seen as a parallel to the topic of *shifts* in translation studies. The concept of a 'shift' in translation is defined by Palumbo (2009: 104) as "a linguistic deviation from the original text, a change introduced in translation with respect to either the syntactic form or the meaning of the Source Text (ST)."[4] The correspondence type hierarchy is not meant to be a new attempt to describe shifts in translation. Firstly, the model is designed with reference to levels of linguistic description (i.e., word forms, syn-

---

[3]The same principles are implicit in the design of the experimental machine translation system PONS, documented in (Dyvik 1990, 1995).

[4]Several researchers have presented taxonomies of the different phenomena involved in translation shifts. The model by Leuven-Zwart (1989; 1990) is frequently cited. For overviews on this topic, see §4 in Chesterman (1997; 2005), and Palumbo (2009: 104-106).

tax, semantics, pragmatics), and it is not from the outset motivated by translation research. Secondly, as will become clear in §3, the type hierarchy model aims not only at describing differences, but also to capture structural parallels, between translationally corresponding units of two languages. Thirdly, there has been a tendency in translation studies to apply the notion of 'shifts' to translation methods, whereas the perspective of the present approach is to measure translational complexity by studying relations between source expressions and their existing translations.

Insofar as the correspondence type hierarchy describes differences in linguistic structure between source and target language expressions, it is thematically connected with research carried out within the field of rule-based MT regarding divergences, and mismatches, between languages. Until the statistical paradigm became dominant, a number of rule-based approaches were developed in order to tackle translation challenges caused by various kinds of differences between languages. An overview of such rule-based translation techniques can be found, e.g., in Trujillo (1999). Barnett et al. (1991) provide a distinction between translation divergences and mismatches which is of relevance to RBMT research. Following Dorr (1990), they describe *translation divergences* as cases where "the same information is conveyed in the source and target texts, but the structures of the sentences are different" (Barnett et al. 1991: 25). Then, referring to Kameyama et al. (1991), they say that *translation mismatches* "occur when there are actually differences in the information that is conveyed" (Barnett et al. 1991: 25). On the background of these two topics, divergences and mismatches, Barnett et al. (1991) argue for the use of interlingual semantic representations in MT development, which is one example of the techniques used in RBMT. Relating to the subject of translation divergences, Dorr et al. (1998: 9-10) present an overview of types of linguistic phenomena that create what they call "mapping problems" in MT; these are basically classes of cases where source and target sentence have different predicate-argument structures, and Dorr et al. (1998: 13-18) discuss these problems in relation to various kinds of system architectures in RBMT.

The distinction between divergences and mismatches, as given by Barnett et al. (1991), is of some relevance to the present study, because it hinges on a notion of 'same information'. If we may assume that this pertains to the information which is encoded linguistically by, respectively, source and target expressions, then cases of translation divergences fall within the domain of computable, or linguistically predictable translation, whereas mismatches represent non-computable, or linguistically non-predictable, translation. E.g., the classes of divergence phenomena described by Dorr (1994) fall within the computable domain of transla-

tion, because they are ascribed to "source-language/target-language distinctions based on lexical-semantic properties" (Dorr 1994: 599), and hence they may be accounted for by information about the two language systems and about how they are translationally interrelated. Her contribution is motivated by the goal of implementing successful MT by means of appropriate techniques. As the present investigation is directed towards measuring translational complexity in existing parallel texts, issues relevant for the implementation of automatic translation will not be discussed further. Moreover, in order to describe specific types of divergences and mismatches, it would be necessary to apply more fine-grained categories than the correspondence types to be presented, and in this study the main focus will be on the distinction between computable and non-computable translation.[5]

3 Methodology

The method applied in this project involves a manual analysis of running parallel texts. In this analysis, translationally corresponding linguistic units, or *string pairs*, are identified and extracted. The chosen units of analysis will be presented in §4.1 Each string pair is analysed according to a classification model, the *correspondence type hierarchy*, which is designed to measure the degree of translational complexity in individual translation tasks.

3.1 The correspondence type hierarchy

In the following, the four main categories of the type hierarchy will be illustrated using examples of sentence pairs taken from a short story by the Norwegian author Bjørg Vik, and its translation into English.

3.1.1 Type 1

The least complex type of translational correspondence is referred to as *type 1*. An example is given in (1), where ((1a) is the source sentence, and (1b) the target sentence:

(1)   a.   Hun har vært en skjønnhet.
           'She has been a beauty.'
      b.   She has been a beauty.

---

[5]Chapter 6 in Thunes (2011) presents a further division of the two most complex correspondence types into subtypes identified by semantic criteria, and these subtypes can be seen as classes of translation divergences and mismatches.

The glossing of (1a) shows that the English target sentence corresponds word-by-word with the source sentence, and this is the characteristic of type 1. That is, in this category, the corresponding strings are pragmatically, semantically, and syntactically equivalent, down to the level of the sequence of word forms. Such correspondences are relatively infrequent in the language pair English-Norwegian.[6]

3.1.2 Type 2

In correspondences of *type 2*, there is also a very close match between the two strings, but there may be some formal differences. Firstly, the sequence of constituents may differ between source and target string; cf. example (2):

(2)   a. Dessuten virket hun overlegen.
           'Also looked she haughty.'
       b. She also looked haughty.

The glossing of (2a) illustrates the word order difference between the two strings. In the Norwegian sentence, there is subject-verb inversion: when a non-subject, such as the adverbial *dessuten*, appears sentence-initially, the verb-second restriction applies in Norwegian. In the English target sentence the subject comes first, and there is no inversion.

Secondly, in type 2 there may be differences in the use of grammatical form words, as shown in example (3):

(3)   a. Leiligheten var ufattelig rotete.
           'Flat.def was unbelievably untidy.'
       b. The flat was unbelievably untidy.

The point in example (3) is that there is no word form in (3a) matching the definite article in (3b), because Norwegian expresses the definite form of nouns by means of a suffix.

The criterion that defines type 2 correspondences is that every lexical word in the source string has a correspondent in the target string of the same lexical category and with the same syntactic function as the source word. This means that in type 2 correspondences, the two strings are pragmatically and semantically equivalent, and equivalent with respect to syntactic functions, even if there is at

---

[6]Table 2 in §4.4 presents the frequencies of the various correspondence types in this study. Similar results were found in Thunes (1998).

least one formal difference that makes the correspondence deviate from word-by-word translation. Type 2 is, like type 1, relatively infrequent in this language pair.

3.1.3 Type 3

In *type 3* correspondences there is, like in types 1 and 2, pragmatic and semantic equivalence between source and target string, but there is not syntactic functional equivalence, because there is at least one structural difference violating equivalence between the two strings with respect to syntactic categories and functions. In the pair of languages English-Norwegian, type 3 seems to be more frequent than each of the two lower types. Type 3 can be illustrated by example (4):

(4) a. Hildegun himlet lidende mot taket og svarte med uforskammet
         høflighet.
         'Hildegun rolled-eyes suffering towards ceiling.def and answered
         with brazen politeness.'
     b. Hildegun rolled her eyes in suffering towards the ceiling and
         answered with brazen politeness.

In this string pair, the correspondence between the Norwegian verb phrase *himlet* and the English expression *rolled her eyes* violates syntactic functional equivalence, because *himle* is an intransitive verb, whereas *rolled her eyes* consists of a transitive verb phrase and a noun phrase functioning as direct object. Also, the Norwegian adverb phrase *lidende* corresponds with the English preposition phrase *in suffering*. Still, these two sentences can be said to correspond semantically.

3.1.4 Type 4

Finally, in *type 4*, the most complex correspondence type, there is no longer semantic equivalence between source and target string. There may be pragmatic equivalence, but not necessarily. In the present study, type 4 has turned out to be very important because it is the most frequent correspondence type in the analysed texts (cf. §4.4).

The defining characteristic of type 4 correspondences is that there is at least one linguistically non-predictable semantic deviation between source and target string. This can be illustrated by example (5):

(5) a. Her kunne de snakke sammen uten å bli ropt inn for å gå i
         melkebutikken eller til bakeren.
         'Here could they talk together without to be called in for to go in
         milk-shop.def or to baker.def'

     b. They could talk here without being called in to go and buy milk or
         bread.

Without going into detail, it may be observed that the semantic difference be-
tween these sentences lies in the correspondence between the substrings *for å
gå i melkebutikken eller til bakeren* and *to go and buy milk or bread*. These ex-
pressions do not denote the same activities, but it is inferrable from background
information about the world that both activities can have the same result, i.e. the
buying of milk or bread.

This illustrates what is involved in a linguistically non-predictable semantic
deviation: the semantic difference between source and target expression — in
the case of example (5), a difference in denotational properties — cannot be pre-
dicted on the basis of the information that is linguistically expressed in the source
string, together with information about source and target languages, and about
their interrelations. This means that in type 4 correspondences, additional in-
formation sources, such as world information, are needed in order to produce
the particular target expression. In such cases, there are normally one or more
alternative translations which *can* be predicted from purely linguistic informa-
tion sources, and which can be semantically equivalent to the original expression.
With respect to (5), a linguistically predictable target expression could be *to go
to the milk shop or to the baker's*. That alternative is denotationally equivalent to
the source expression, but it does not necessarily exhibit other properties that a
translator may want to choose in a target text.

3.2 Some aspects of the classification model

The examples (1)–(5) show that the correspondence type hierarchy, as a classifi-
cation model, reflects a gradual increase in linguistic divergence between source
and target string, and the analysis of translational correspondences is based on
the assumption that this increase is correlated with an increase in the degree of
translational complexity. That is, a larger amount of information, and a greater
processing effort, is required in order to solve translation tasks in correspon-
dences of the higher types than in the lower types.

Each correspondence type covers a class of translation tasks, and in the type
hierarchy, the four classes are distinguished from each other on the basis of the

types and amount of information necessary for solving translation tasks within each class. These matters are described in detail for each correspondence type in Thunes (2011), along with discussions of the accessibility of necessary information sources, and of required processing effort, within each type.

On the scale of translational complexity defined by the type hierarchy, the division between predictable and non-predictable translation is drawn between the types 3 and 4. This means that correspondences of types 1, 2, and 3 together constitute the domain of linguistically predictable, or computable, translations, whereas type 4 correspondences belong to the non-predictable, or non-computable, domain, where semantic equivalence is not fulfilled.

A clear parallel to the increasing degree of complexity in the type hierarchy is found in Vinay and Darbelnet's set of seven translation procedures, which are presented "in increasing order of difficulty", ranging from the simplest method of translation to the most complex.[7] Although this is an interesting similarity, the present classification model is not related to Vinay & Darbelnet's categorisation of methods.[8]

Moreover, the classification of correspondences involves no evaluation of translational quality as, for instance, in terms of the model by House (1997). Among the empirical data there are occasional instances of unsuccessful translations, but translational quality is by itself no element in the classification of correspondences. Moreover, our notion of translational complexity, being based on information sources for translation, is in principle independent of grammatical complexity, and of factors that may influence the ease or difficulty with which the translator comprehends the source text.[9] Translational complexity is also distinct from the notion of linguistic complexity, as defined, e.g., by Dahl (2004).

3.3 Predictability and information sources for translation

In the present approach, the distinction between computable and non-computable translation is the same as the dichotomy between linguistically predictable and non-predictable translation (cf.§1.1), and the distinction relies further on a typology of information sources for translation, presented in Thunes (2011: 87–106). In relation to the computability issue, the most important distinctions drawn in that typology are, firstly, the division between linguistic and

---

[7]The quotation is taken from Venuti (2000: 92), where an overview of the seven procedures is presented. Pages 31–42 of Vinay & Darbelnet (1995) are reprinted in Venuti (2000: 84-93).

[8]Cf. the comments in §1.1 and §2 on the product-orientation of the present approach, and on how this study is related to translation shifts.

[9]Grammatical complexity in relation to translation is discussed by Izquierdo & Borillo (2000).

extra-linguistic information, and, secondly, the borderline between information coded inside the source language expression, and information available from the context of that expression.

Within the *linguistic information sources* for translation there is, firstly, the information supporting the translator's knowledge of source and target language systems and their interrelations. Secondly, these sources include the information that is linguistically encoded in the source expression. This covers information about the situation type described by the source text, information about the linguistic structure of the source expression, as well as information about relations of reference holding between expressions in the source text and extra-linguistic entities. The latter is derivable from the source language expression as it is interpreted in a specific context. Thirdly, the linguistic sources also include information available in the linguistic context of the source string.

The *extra-linguistic information sources* for translation comprise general background information about the world, information about particular technical domains, information about textual norms, and information derivable from previous translation training and practice. They also cover information about the utterance situation of the source text, and about the translation situation. These types may include information about the sender, about the purpose(s) of original and translation, about temporal and geographical location, etc. Another extra-linguistic information source may be information derived by applying different kinds of background information in common-sense reasoning about facts described by the SL text.

The division between linguistic and extra-linguistic information can be briefly illustrated with reference to examples (1)–(5) in §3.1.1–§4. In examples (1)–(4), which instantiate correspondence types 1–3, each target sentence can be predicted from the source sentence by means of linguistic information sources alone. That is, the translations can be computed on the basis of the information that is linguistically encoded in the source sentences, together with information about source and target language systems, and their interrelations. Then, in example (5), linguistic information sources are not sufficient in order to produce the target sentence, as background information about the world is also required.[10]

If we consider a language system to be a structure containing a finite set of components which may be combined in a limited number of possible ways, then it may be argued that information about a language system in principle constitutes a finite domain. The extra-linguistic world, on the other hand, is unlimited,

---

[10]The dichotomy between linguistic and extra-linguistic information is discussed further in Thunes (2011: 90–102).

and hence information about it can be regarded as a nonfinite domain. Then, a strategy for separating linguistic from extra-linguistic information is to delimit the given language system, and, in line with Dyvik (2003: 9), the distinction between the linguistic and the extra-linguistic is thus related to the way in which language systems are conceptually individuated. This, in turn, will be influenced by the purpose for which the language description is meant to be applied, and by empirical facts about language use; cf. Thunes (2011: 93).

The information coded in a specific source language expression is necessarily finite, delimited by the expression itself. The information available from the surrounding context is in principle unbounded, although there is in practice a limit on how much context that will be considered by the translator when producing a target expression.

Accordingly, a target expression that can be predicted from the information coded linguistically in the source string, together with information about source and target language systems and their interrelations, can be regarded as computable because there is a finite search space which contains the information needed to produce that target expression. Likewise, if a translation is non-computable, then some information falling outside the finite, linguistic domain is required in order to create that particular target expression.

For a given source expression, there is normally a number of possible translations, and the appropriateness of each alternative is typically context-dependent. A subset of all the possible alternatives in the target language will be computable, or linguistically predictable translations, determined by information about the interrelations between the two languages. In the present study, the classification of translational correspondences amounts to deciding, for each target string, whether it belongs to the set of predictable translations of the given source string, or not. If it does not, the correspondence is non-computable, and of type 4. If it does, the string pair is computable, and of type 1, 2, or 3, depending on the degree of linguistic convergence between source and target expression.

This analytical approach relies on a certain understanding of linguistic approaches to MT: automatic translation is seen as possible to the extent that the translation system has access to information about source and target languages and their interrelations, and from those information sources only linguistically predictable translations can be generated by the system. There is a principled difference between this and human translation, because the human translator chooses a predictable translation only when it appears to be the most appropriate choice also on the basis of information falling outside the finite, linguistic domain.

4 Empirical investigation

The implementation of the present methodology involves manual extraction and classification of string pairs from parallel texts. The application of the type hierarchy requires a human, bilingually competent analyst, since the classification of the compiled correspondences demands a careful linguistic analysis of each string pair.

The assignment of correspondence type to individual string pairs works like an elimination procedure where we start by testing for the lowest correspondence type and then move upwards in the hierarchy if the test fails. This may seem a fairly straightforward task, but not in every case. In particular, it can be difficult to distinguish between instances of types 3 and 4, since that may involve fine-grained semantic analyses.

4.1 Units of analysis

A limited set of syntactic units have been chosen as units of analysis, and the selection of units is influenced by the wish to make this study of translational complexity relevant to the field of machine translation. It has been an aim to find a way of segmenting text material that would be suitable for automatic translation regardless of specific algorithms for implementation. Considering the linguistic approaches to automatic translation, MT systems typically operate sentence by sentence, and hence *the finite clause* is chosen as the basic unit of analysis in this investigation. Another point motivating the choice is that in order to be of any use, an MT system must handle syntactic units at least as complex as those of the sentence level.

In this connection, 'finite clause' is understood simply as a syntactic unit containing a finite verb as its central element. Thus, occurrences of finite verbs are in practice the basis for the identification of analysis units. Whenever a word form of this category is encountered, the syntactic unit in which it fills the function of main or auxiliary predicate is identified as a unit of analysis. This means that matrix sentences and finite subclauses are typically recorded as units of analysis. Also, lexical phrases with one or more finite clauses as syntactic complement (cf. (6) in §5.2) constitute another major syntactic type among the recorded data. In such cases the finite clause is not identified as an independent unit, because the entire phrase is normally a more natural unit of translation than the syntactic complement in isolation.

The parallel texts are analysed from beginning to end. The human annotator goes through the texts in parallel in order to identify pairs of translationally

related units. Notably, string pairs are extracted also when only one of the two strings is a syntactic unit satisfying the criteria by which units of analysis are identified. This is necessary because finite constructions may be translationally related to nonfinite constructions, and such correspondences are frequent in the language pair English-Norwegian. Once a unit of analysis, and its translational correspondent, are identified, the string pair is recorded and a correspondence type assigned to it. The data are stored electronically by means of the software tool Text Pair Mapper, described in Dyvik (1993).

As syntactically dependent constructions like finite subclauses occur as units of analysis, the data include nested correspondences where a superordinate string pair contains one or more embedded string pairs. E.g., if a finite subclause is embedded in a matrix sentence, as in *When he came, we could leave* (Norwegian: *Da han kom, kunne vi dra*), then two string pairs are extracted. One is the subclause and its match in the parallel text: [*When he came,*] – [*Da han kom*]; the other is the matrix sentence and its correspondent: [[CP] *we could leave*]. – [[CP] *kunne vi dra*]. In the superordinate string pair, the embedded correspondence is treated as a pair of opaque items, represented by their syntactic categories.[11]

4.2 The texts

In this study, the data are recorded from a selected set of English-Norwegian parallel texts. The texts were written during the years 1979–1996, and all translations have been produced manually. The corpus covers both directions of translation, and it includes two text types, fiction and law texts. Comparable amounts of data have been compiled for each of the text types and directions of translation. Table 1 gives an overview of text type, direction of translation, and numbers of running words for each of the text pairs that have been investigated.

4.2.1 Degree of restrictedness

In the present investigation, law texts are chosen as a representative of restricted text types, and fiction as an example of a relatively unrestricted type. The difference in restrictedness between the two text types is a direct reflection of a basic opposition between the language of the law and that of fiction: the former is strictly norm-governed in ways that the latter is not. In law-regulated societies the law is nothing less than the highest power, and this gives law texts their authority. Because of the optimally authoritative status of a law text, its production

---

[11]In the present analysis, the category label *CP* represents finite subclauses; cf. Thunes (2011: 201).

Table 1: An overview of the analysed text pairs with respect to text type, direction of translation,and numbers of running words.

Authors and texts	Text type	Source	Target	No. of running words
Agreement on the European Economic Area, Articles 1–99	law text	Eng.		9202
Avtale om Det europeiske økonomiske samarbeidsområde artiklene 1–99			Nor.	8015
Lov om petroleumsvirksomhet §§1–65	law text	Nor.		7929
Act relating to petroleum activities, Sections 1–65			Eng.	9647
André Brink	fiction			
The Wall of the Plague		Eng.		4021
Pestens mur			Nor.	4230
Doris Lessing	fiction			
The Good Terrorist		Eng.		4008
Den gode terroristen			Nor.	4652
Erik Fosnes Hansen	fiction			
Salme ved reisens slutt		Nor.		4022
Psalm at Journey's End			Eng.	4395
Bjørg Vik	fiction			
En håndfull lengsel		Nor.		4010
Out of Season and Other Stories			Eng.	4550
Total				68681

as well as its interpretation are strongly governed by the intersubjective norms of the legal domain of society; cf., e.g., Bowers (1989: 53-54), and Cao (2007: 13-14). According to Bhatia (2010: 46-47), the primary concern in law writing is "loyalty to legislative intentions", and he describes four different norms of law writing: clarity of expression (i.e. avoiding vagueness), precision (by using as few words as possible), unambiguity, and all-inclusiveness (i.e. specifying adequately the scope of application of the law text) (2010: 38-39).

Fiction texts, then, are, like any kind of language use, subject to the linguistic norms of the language community, and there are norms of literary language use that shape the characteristics of various kinds of styles and genres.[12] Still, fiction texts are in no way as norm-governed as law texts are, and although literary norms, too, have intersubjective existence, they are not institutionalised like legal norms. As a parallel to the authority of law texts, fiction texts can acquire high status if they are particularly successful. In such cases, the status of the fiction text is determined, firstly, by the creative ability of the author to express a story, and, secondly, by the capacity of that story to create great experiences in the minds of the readers. The subjective factors attributed to the sender and recipient of a fiction text are quite different from the institutionalised norms controlling the writing and interpretation of law texts. The production of a fiction text is governed by the individual choices of the author, which may include norm violations, and its reception is determined by the subjective experiences of the readers. This is in sharp contrast to the norms of law texts, which are determined by the collective purpose of regulating society.

The text-typological differences between law and fiction are evident in relation to translation. Since the meaning of a law text expresses legal content, the translation must preserve the meaning of the original as far as possible, given differences in semantic structure between the two languages. When fiction is translated, there may be other properties than the linguistic meaning of the source text that are necessary to recreate in the target text. In particular, it will be important to preserve the literary properties of the original, and hence the choice of target expressions can be motivated by a range of other factors than the semantic content of the source text. This point is further commented on in §5 with reference to observations of semantic deviations in the empirical data of the present study.

---

[12]The kind of norms that shapes the linguistic characteristics of literary styles is described by Leech & Short (2007: 41-44) as *relative norms*.

4.2.2 Textual features

Various kinds of linguistic effects of the difference in restrictedness between the investigated text types are discussed in Thunes (2011: 279–288). The principal consequence of this opposition is that there is a greater degree of structural diversity in the fiction texts than in the law texts, and this is evident from a range of features that can be observed in the selected texts.

The investigated law texts exhibit several features which are characteristic of this text type. They contain sets of sequentially numbered sections, or articles, and are written in a formal, impersonal style, with frequent use of long, complex sentences. Mattila (2006: 98), citing Laurén (1993: 74), observes that "[s]entences in legal language are longer than those of other languages for special purposes and they contain more subordinate clauses." The texts are repetitive in the sense that specific expressions are recurrent (e.g., *with a view to*, *without prejudice to*). Other characteristics are heavy constituents, enumerative listing, complex coordination, no occurrences of first and second person pronouns, and numerous instances of nonfinite constructions, especially in the English texts. Another salient feature is the high frequency of headings, normally realised as noun phrases, such as *Article 1*. The texts contain a limited inventory of types of sentences and syntactic constructions, and short, syntactically simple sentences are infrequent.

The analysed fiction texts are extracts of novels, except for the text by Vik (1979), which is taken from a short story (cf. Table 1). Each extract runs from the beginning of the narrative, and none of them is a complete text. The selected fiction texts are stories evolving around a certain protagonist and other characters, and passages of dialogue are found in all of them. In comparison to law text, narrative fiction can rightly be described as unrestricted, at least in terms of the inventory of syntactic constructions that may occur. Narrative fiction texts may comprise all kinds of sentence types: simple as well as complex, declarative, interrogative, and imperative sentences.[13] Furthermore, literary texts can include direct speech and passages of other text types, which may add to the structural diversity. Moreover, as discussed in Thunes (2011: 283–284), the analysed fiction texts exhibit a larger variety of speech acts than the law texts do, and this is clearly linked with the greater degree of structural diversity in the fiction texts.

---

[13]Cf. Ochs (1997: 185–189) on the diversity of narratives. On the narrative in general, see Abbott (2002), and Toolan (2001).

4.3 Measuring translational complexity

In order to measure the degree of translational complexity in pieces of parallell texts, the classification model must be applied to running texts, without omitting any parts of them. Then, the distribution of the four correspondence types within a set of data provides a measurement of the degree of translational complexity in the parallel texts that the data are extracted from.

Calculating the distribution of correspondence types brings attention to the difference between counting the frequencies of string pairs of each type and measuring the length of text covered by each category. The reason for this is that, in the given language pair, the two least complex types (1–2) normally occur in pairs of short and syntactically simple strings of words, whereas pairs of longer and more complex strings tend to be of the two higher types (3–4). Thus, types 1 and 2 would appear as covering an unproportionally large amount of the analysed texts if the distribution of the main correspondence types would be presented merely on the basis of the numbers of string pairs (cf. Table 2 in §4.4).

Hence, the proportions of text covered by the different correspondence types will be discussed in terms of the lengths of, respectively, source and target text. More precisely, the proportions are measured by means of string lengths, i.e. by calculating the number of word forms covered by each correspondence type. The length of a recorded translational unit equals its number of word forms, and in the case of nested correspondences, the word forms in embedded strings are counted only once. That is, if a recorded unit contains any embedded strings, then each embedded unit is treated as an opaque unit in the superordinate string. The length of the matrix unit is counted as its number of non-opaque word forms, and a subordinate unit adds only 1 to the length of the superordinate string.

The most important aspect shown by the complexity measurements of this study is the division between computable and non-computable correspondences, i.e. how large is the proportion of the analysed texts covered by, on the one hand, string pairs of types 1, 2, and 3, and, on the other hand, string pairs of type 4. This division is meant to show to what extent it can be expected that an ideal, rule-based MT system could simulate the given translations, if provided with a full description of the two languages and their interrelations. Notably, this is not an estimate of how much of the given source texts that could be given *some kind of* linguistically predictable translation. Since English and Norwegian both belong to the Germanic language family, and are used in language communities which are, in cultural terms, not very far apart, the recorded data include probably only very few source expressions which have no linguistically predictable

translation.[14] It should be emphasised, then, that this study tries to measure the proportion of predictable, and hence computable, translation within the specific, human-created target texts that already have been produced (cf. §3.3).

4.4 The results

Since the present investigation is based on hand-coded material, the data are of a relatively modest quantity (about 68 000 words), and it will remain a mere speculation whether the distribution of correspondence types across the total set of data may reflect the general degree of complexity in the translational relation between English and Norwegian, as instantiated in actual, human translation activity. The limited size of the compiled data prevents the detection of statistically significant results, and only tendencies may be observed within the recorded material. Hence, it is not possible to generalise about the degree of translational complexity in relation to the given language pair, the two directions of translation within this pair, or to the investigated text types. Still, on the basis of the recorded data, the results provide tentative answers to the research questions posed in §1. After a brief presentation here, the results will be further discussed in §5 with subsections.

Concerning the automatisation issue, Table 2 shows the complexity measurement across the entire collection of correspondences. By calculating the average values of the percentages given for, respectively, source and target text lengths, we find that more than half of the data are included in non-computable correspondences: string pairs of type 4 constitute 55.2% of the compiled data, whereas the computable types 1, 2, and 3 together cover as little as 44.8% of all recorded string pairs. On the basis of this result, the conclusion is that with perfect information about source and target languages, an idealised rule-based MT system could have simulated less than half of the identified correspondences.

Further, Table 2 shows that within the subset of computable correspondences, type 3 constitutes a large majority of the data. Thus, types 1 and 2 together cover a very modest proportion of the analysed texts (on average 10.0% across all data), and this strengthens the point made in §4.3 that the most important aspect shown by the data is the division between, on the one hand, types 1–3 and, on the other hand, type 4. Because types 1 and 2 are so infrequent, the distinction between computable and non-computable correspondences appears to be the most informative indicator of translational complexity, as far as the language pair English-Norwegian is concerned.

---

[14] An example could be the Norwegian noun *skiføre*, found in Vik's text (1979a). This word has no match in English, and needs to be translated by a paraphrase, such as *conditions for skiing*.

Table 2: The global distribution of correspondence types in the investi-
gated texts.

Total results, all text pairs	Type 1	Type 2	Type 3	Type 4	All types
Number of string pairs	601	272	1 347	2 219	4 439
Percentage of string pairs	*13.5*	*6.1*	*30.4*	*50.0*	*100.0*
Source text length (word forms)	1 906	1 642	12 179	19 263	34 990
Percentage of source text length	*5.4*	*4.7*	*34.8*	*55.1*	*100.0*
Target text length (word forms)	1 926	1 741	12 940	20 547	37 154
Percentage of target text length	*5.2*	*4.7*	*34.8*	*55.3*	*100.0*

With respect to the text type issue, the results are summed up in table 3, which
shows that the proportion of computable correspondences is on average 50.2%
in the law data, and 39.6% in fiction. However, it is pointed out in Thunes (2011:
275) that these results cannot be seen as indicative of the general complexity of
translating, respectively, law text and fiction between English and Norwegian.
The results in Table 3 merely show that the degree of complexity is, on average,
lower in the selected pairs of law texts than in those of fiction.

Table 3: Differences in translational complexity between the two text
types.

Proportions of...	in law text	in fiction	in all data
computable translational correspondences (types 1, 2, 3)	50.2%	39.6%	44.8%
non-computable translational correspondences (type 4)	49.8%	60.4%	55.2%

Moreover, the results do not indicate that while the analysed fiction texts ap-
pear as clearly unsuitable for automatic translation, the law texts appear as suit-
able. Across the investigated material, the degree of translational complexity is
found to be so high that fully automatic translation does not seem to be a fruitful
option for any of the analysed text pairs, if the aim is to produce output identical
to the human-created target texts of the analysed data. Furthermore, as explained
in Thunes (2011: 275), the lower degree of average complexity in the chosen law
text pairs is primarily due to the relatively low complexity measured in the law
text translated from Norwegian into English (60.9% computable translation; cf.

Thunes (2011: 291). In the other pair of law texts, the degree of complexity is higher, and, in fact, quite similar to the average found across the four pairs of fiction texts (39.6% computable translation; cf. Table 3).

5 Discussion

In relation to the automatisation issue, the results are rather pessimistic, especially considering the fact that automatic translation tools are actually used, in particular for non-literary text types, and this is so because they do reduce the workload of manual translation. Better performance may be expected by MT systems developed for restricted domains, or subject areas, and it is also likely that some of the non-computable correspondences among the recorded data could have been maintained by translation memories.[15] The latter is highly relevant for law texts, which tend to be repetitive (cf. §4.2.2). §5.1 and §5.2 provide further comments on the automatisation issue.

Concerning the text type issue, it is an expected result to find a lower degree of translational complexity in law texts than in fiction texts. Chapter 6 in Thunes (2011) provides discussions of several kinds of recurrent semantic deviations between translationally corresponding units, and, in general, these phenomena constitute the primary factor contributing to the frequency of the most complex correspondence type. Although cases of type 4 are not infrequent within the law data, instances of semantic deviations are far more common among the fiction data than among those compiled from law text. This is in line with the high degree of restrictedness in the law texts (cf. §4.2.1). In particular, since legal translation is strongly governed by the norm of preserving the informational content of the original in the target text, the abundance of semantic deviations found in the fiction target texts would be simply unacceptable in the domain of law translation.

Given the dominance of statistical machine translation, it may appear surprising that this study assumes the traditional, linguistic approach to MT, where translations are computed on the basis of formal descriptions of source and target language systems and their interrelations. However, in recent years the general view has been formed that there is a limit to how far the purely statistical methods can reach in terms of translation quality, and for more than a decade research efforts have been put into hybrid approaches where statistical techniques

---

[15]A *translation memory* is defined by Palumbo (2009: 127–128) as "[a]n electronic database containing translated texts stored together with their originals," and the texts "are normally segmented into units one sentence long."

are combined with some kind of semantic and/or syntactic processing. If a certain level of quality is wanted, it seems unlikely that automatic translation can do without linguistic information, especially in the light of the pervasive ambiguity of natural language expressions.

Still, the general issue of computability, or linguistic predictability, which is behind the present approach should in principle also apply to statistical machine translation, because SMT, too, depends on the accessibility of relevant and sufficient information within the texts themselves in order to predict correct target expressions from available translational correspondences.

5.1 Human translations as a gold standard

With respect to the automatisation issue, translations produced by humans have been used as a gold standard for MT in this study. In relation to this, it is a point that the analysed texts provide a problematic norm for automatic translation. Since it is generally accepted that the use of machine translation requires post-editing to secure the quality of the final product, the human-created target texts represent an ideal for the end result, and not for the raw output of an MT application. The chosen norm is probably an unrealistic, and perhaps also unfair, goal for MT development, especially since high-quality translation without post-editing, or revision, is uncommon also when the translator is human. Still, manually produced target texts have been used as a standard because evaluating the products of real systems has not been an objective, and because the complexity measurements in this study aim at showing to what extent we might assume that an ideal, rule-based system could simulate the given translations.

5.2 Minimally non-computable correspondences

In order to discuss further whether it would be fruitful to apply automatic translation to the selected texts, it is interesting to consider the workload potentially involved in editing possible machine output. For this purpose, we can assume that an MT system would generate only linguistically predictable translations for the analysed source texts. This means that the recorded type 4 correspondences represent cases where the machine would produce target expressions conforming with the characteristics of one of the lower correspondence types, or possibly not generate linguistically well-formed output at all. At any rate, post-editing would be required in order to reach the gold standard represented by the human-created translations.

Of relevance here is the question whether string pairs identified as type 4 in the present study have been classified as such because of only one, or few, semantic deviations between source and target units. That is, if the semantic difference between two corresponding strings is small, then the major part of the correspondence would involve linguistically predictable translation, and it might be unproblematic for a post-editor to correct that subpart of the machine output which does not meet the standard. If post-editing amounts to simple corrections of linguistic errors that are few and easy to spot, then what Jurafsky & Martin (2009: 931), describe as the *edit cost* of post-editing would be low, and the *editing distance* between the machine output and the standard could be small, and automatic translation might be useful.[16] On the other hand, if there are many errors in the output, and, if the revision also requires syntactic and/or semantic reorganisation of the automatically generated sentences, and maybe even careful considerations of the appropriateness of various target alternatives, then the editing distance is large, and it is perhaps more cost effective to do a fully manual translation.

As mentioned in §5, a set of recurrent semantic deviations between translationally corresponding units have been identified among the recorded data, and these phenomena are likely to represent challenges that the post-editor will be faced with, i.e. types of properties that should be observed in the translation, but which cannot be predicted from the source expression without access to contextual information, and/or various kinds of extra-linguistic information. It is of significance to the question of potential edit cost that the editing distance between , on the one hand, a predictable, machine-generated translation and, on the other hand, a human-created target string with multiple semantic deviations in relation to the original will be considerably greater than the distance between a predictable translation and a target expression exhibiting only a minimal semantic difference in relation to the source string.

Thus, non-computable correspondences with only one minimal semantic deviation between source and target string are of particular interest to the question of potential edit cost. Such cases may be described as *minimally non-computable*, and in correspondences of this kind it would probably be easy to revise an automatically generated target expression to the standard of manual translation. An example can be taken from the Norwegian *Act relating to petroleum activities*.

---

[16]The term *editing distance* is borrowed from information theory. According to Jurafsky & Martin (2009: 108), "[t]he minimum edit distance between two strings is the minimum number of editing operations (insertion, deletion, substitution) needed to transform one string into another."

The noun phrase given in ((6)a) contains a relative clause, and is translated into the expression shown in ((6)b):

(6) a. de områder som er nevnt i tillatelsen
         'the areas which are mentioned in license.def'

    b. the areas mentioned in the licence

The only semantic deviation in this string pair is the presence vs. absence of grammatically expressed temporal information, and because of this, example (6) is a type 4 correspondence. Here it can be assumed that a rule-based translation system would produce the semantically equivalent target expression *the areas which are mentioned in the licence*, and a human post-editor might easily choose the nonfinite alternative because he or she would know that that would be stylistically more appropriate in a law text.

In a metric for evaluating MT output, Specia (2011: 75) distinguishes between four degrees of quality, ranging from the lowest one where complete retranslation is required, to the highest degree where the output is a fully acceptable translation. Intermediate degrees on this scale are cases where the translation is not very good, but post-editing is less demanding than retranslation, and cases where very little editing is needed. Given the assumption that minimally noncomputable correspondences represent translation tasks where the editing cost would be very low, there is a close affinity between this category and Specia's second highest degree of quality.

The distribution of minimally non-computable correspondences among the recorded data again puts focus on the text type issue, because such cases are far more frequent in the law texts than in the fiction texts. Within the law data, as much as 45.7% of the correspondences classified as type 4 are minimally noncomputable, whereas among the fiction data, only 10.5% of the compiled type 4 correspondences are minimal ones. This primarily reflects the fact that because law text is strongly norm-governed in a way that fiction text is not (cf. §4.2.1), semantic deviations between translationally corresponding units are far less frequent in the former than in the latter. Moreover, it shows that the potential edit cost required by automatic translation would be considerably lower in the law texts than in the fiction texts.

5.3 Conclusions and a possible extension

On the basis of the data recorded in this study, the investigated pairs of law texts are tentatively regarded as representing a text type where machine translation

may be helpful, if the effort required by post-editing is smaller than that of manual translation. In the case of the fiction texts, it seems clear that post-editing of automatically generated translations would be laborious and not cost effective.

The careful optimism in relation to the automatisation of law text translation is not only inspired by the findings of the present investigation, but also by the recent emergence of a research field combining insights and methods from artificial intelligence, human language technology, the law, legal informatics, and studies of legal language. E.g., under the heading *Semantic Processing of Legal Texts*, Francesconi et al. (2010) have compiled a set of contributions dealing with topics such as information extraction from legal texts, the construction of legal knowledge resources, semantic indexing, summarisation, and translation evaluation for the legal domain. Furthermore, Johnsen (2010), and Johnsen & Berre (2010) discuss the semantic modelling of law text with reference to Norwegian. Contributions like these indicate that there is progress in relation to the development of automatic analysis of law text. Moreover, since the language of law is highly specialised and norm-controlled, it is, in its own right, of interest to the field of language technology as a testing ground for applications developed for the processing of natural language, translation included.

Then, I will suggest that the correspondence type hierarchy has a potential as a diagnostic tool for the feasibility of linguistics-based machine translation in relation to specific text types. That is, by applying the method to limited selections of parallel texts of the same type, it would be possible to estimate to what extent the target text could be generated automatically. If the proportion of assumed computable correspondences would exceed a chosen threshold, it might be worthwhile to tune an MT system for the given language pair to the text type in question, for instance by developing lexicon modules covering the relevant subject domain.

Moreover, since the feasibility of MT for a given text type is determined also by potential edit cost, it would be fruitful to extend the classification model by integrating a fifth correspondence type to be assigned to the minimally non-computable string pairs. If such a fifth category could be implemented in the software used for recording translational correspondences, it would be possible to calculate automatically the proportion of minimally non-computable correspondences in terms of string length within each text pair. Such estimates could say something about the potential edit cost required by automatic translation.

Finally, we may recall that adaptations of the methodology of the present study have been put to use in several works within the field of contrastive linguistics (cf. §2). Moreover, as discussed in Thunes (2011: 446–447) the data analysed

for the purposes of this investigation do not only say something about translational complexity; they also shed some light on how the language systems of English and Norwegian are interrelated, and they reveal aspects of the relation between source and target texts in the analysed corpus. Thus, the present project illustrates that the different fields of machine translation, contrastive language research, and translation studies have an important common denominator in the analysis of translational correspondences.

6 Acknowledgments

References

Abbott, H. Porter. 2002. *The Cambridge introduction to narrative*. Cambridge: Cambridge University Press.

Azevedo, Flávia. 2012. *The problem of codifying linguistic knowledge in two translations of shakespeare's sonnets: A corpus-based study*. Florianópolis: Federal University of Santa Catarina dissertation. Doctoral dissertation.

Barnett, James, Inderjeet Mani, Elaine Rich, Chinatsu Aone, Kevin Knight & Juan Carlos Marinez. 1991. Capturing language-specific semantic distinctions in Interlingua-Based MT. In *Proceedings of Machine Translation Summit III*, 25–32. Washington D.C.

Bhatia, Vijay K. 2010. Specification in legislative writing: Accessibility, transparency, power and control. In Malcolm Coulthard & Alison Johnson (eds.), *The Routledge handbook of forensic linguistics*, 37–50. London: Routledge.

Bowers, Frederick. 1989. *Linguistic aspects of legislative expression*. Vancouver: University of British Columbia Press.

Cao, Deborah. 2007. *Translating law: Topics in translation 33.* Clevedon: Multilingual Matters Ltd.

Chesterman, Andrew. 1997. *Memes of translation: The spread of ideas in translation theory* (Benjamins Translation Library 22). Amsterdam: Benjamins.

Chesterman, Andrew. 2005. Problems with strategies. In Krisztina Károly & Ágota Fóris (eds.), *New trends in translation studies: In honour of Kinga Klaudy*, 17–28. Budapest: Akadémiai Kiadó.

Cyrus, Lea. 2006. Building a resource for studying translation shifts. In *Proceedings of the Fifth International Conference on Linguistic Resources and Evaluation (LREC-2006)*, 1240–1245. Genoa.

Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity* (Studies in Language Companion Series 71). Amsterdam: Benjamins.

Dorr, Bonnie J. 1990. Solving thematic divergences in machine translation. In *Proceedings of the 28th Annual Meeting of the ACL*, 127–134. Pittsburgh.

Dorr, Bonnie J. 1994. Machine translation divergences: A formal description and proposed solution. *Computational Linguistics* 20(4). 597–633.

Dorr, Bonnie J., Pamela W. Jordan & John W. Benoit. 1998. *A Survey of Current Paradigms in Machine Translation.* Tech. rep. University of Maryland, College Park. Technical report.

Dyvik, Helge. 1990. *The PONS project: Features of a translation system* (Skriftserie fra Institutt for fonetikk og lingvistikk 39, B). Bergen: University of Bergen.

Dyvik, Helge. 1993. *Text Pair Mapper.* Unpublished manuscript.

Dyvik, Helge. 1995. Exploiting structural similarities in machine translation. *Computers and the Humanities* 28(4/5). 225–234.

Dyvik, Helge. 1998. A translational basis for semantics. In Stig Johansson & Signe Oksefjell (eds.), *Corpora and cross-linguistic research: Theory, method, and case studies* (Language and Computers: Studies in Practical Linguistics 24), 51–86. Amsterdam: Rodopi.

Dyvik, Helge. 2003. *Translations as a Semantic Knowledge Source. Unpublished manuscript.* http://www.hf.uib.no/i/LiLi/SLF/ans/Dyvik/transknow.pdf. (last accessed on 22 June 2012).

Elgemark, Anna. 2017. *To the very end: A study of N-Rhemes in English and Swedish translations.* University of Gothenburg dissertation. Doctoral dissertation.

Francesconi, Enrico, Simonetta Montemagni, Wim Peters & Daniela Tiscornia. 2010. *Semantic processing of legal texts: Where the language of law meets the law of language.* Vol. 6036 (Lecture Notes in Artificial Intelligence). Berlin: Springer.

Hasselgård, Hilde. 1996. Some methodological issues in a contrastive study of word order in English and Norwegian. In Karin Aijmer, Bengt Aijmer & Mats Johansson (eds.), *Languages in contrast. Papers from a Symposium on Text-based Cross-linguistic Studies. Lund 4–5 March 1994* (Lund Studies in English 88), 113–126. Lund: Lund University Press.

Hasselgård, Hilde. 1998. Thematic structure in translation between English and Norwegian. In Stig Johansson & Signe Oksefjell (eds.), *Corpora and cross-linguistic research: Theory, method, and case studies* (Language and Computers: Studies in Practical Linguistics 24), 145–167. Amsterdam: Rodopi.

House, Juliane. 1997. *Translation quality assessment: A model revisited* (Tübinger Beiträge zur Linguistik 410). Tübingen: Gunter Narr.

Izquierdo, Isabel García & Josep Marco Borillo. 2000. The degree of grammatical complexity in literary texts as a translation problem. In Allison Beeby, Doris Ensinger & Marisa Presas (eds.), *Investigating translation. Selected papers from the 4th International Congress on Translation, Barcelona, 1998* (Benjamins Translation Library 32), 65–74. Amsterdam: Benjamins.

Johnsen, Åshild. 2010. *Forstå det den som kan. Semantisk modellering av juridisk regelverk med bruk av SBVR – en brobygger mellom jus og IT* MA thesis.

Johnsen, Åshild & Arne-Jørgen Berre. 2010. A bridge between legislator and technologist – Formalization in SBVR for improved quality and understanding of legal rules. In Thomas Eiter, Adil El Ghali, Sergio Fernàndez, Stijn Heymans, Thomas Krennwallner & François Lévy (eds.), *Proceedings of BuRO 2010: 1st International Workshop on Business Models, Business Rules and Ontologies*, 29–39. Brixen.

Jurafsky, Daniel & James H. Martin. 2009. *Speech and language processing. An introduction to natural language processing, computational linguistics, and speech recognition.* Upper Saddle River: Pearson Education. Second edition.

Kameyama, Megumi, Ryo Ochitani & Stanley Peters. 1991. Resolving translation mismatches with information flow. In *Proceedings of the 29th Annual Meeting of the ACL*, 193–200. Berkeley.

Laurén, Christer. 1993. *Fackspråk: Form, innehåll, funktion.* Lund: Studentlitteratur.

Leech, Geoffrey & Mick Short. 2007. *Style in fiction: A linguistic introduction to English fictional prose* (English Language Series). Harlow: Pearson Education Limited. Second edition.

Leuven-Zwart, Kitty M. van. 1989. Translation and original: Similarities and dissimilarities, i. *Target* 1(2). 151–181.

Leuven-Zwart, Kitty M. van. 1990. Translation and original: Similarities and dis-
similarities, ii. *Target* 2(1). 69–95.

Macken, Lieve. 2010. *Sub-sentential alignment of translational correspondences.*
Antwerp dissertation.

Mattila, Heikki E. S. 2006. *Comparative legal linguistics.* London: Ashgate.

Merkel, Magnus. 1999. *Understanding and enhancing translation by parallell text
processing.* Dissertation No. 607.

Ochs, Elinor. 1997. Narrative. In Teun A. van Dijk (ed.), *Discourse as structure and
process: Discourse studies: A multidisciplinary introduction 1*, 185–207. London,
Thousand Oaks, & New Dehli: Sage Publications.

Palumbo, Giuseppe. 2009. *Key terms in translation studies.* London & New York:
Continuum.

Silva, Norma Andrade da. 2008. *Análise da tradução do item lexical* evidence *para
o português com base em um corpus jurídico.* Florianópolis: Federal University
of Santa Catarina MA thesis.

Specia, Lucia. 2011. Exploiting objective annotations for measuring translation
post-editing effort. In Mikel L. Forcada, Heidi Depraetere & Vincent Vandegh-
inste (eds.), *Proceedings of the 15th Conference of the European Association for
Machine Translation*, 73–80. Leuven, Belgium.

Thunes, Martha. 1998. Classifying translational correspondences. In Stig Johans-
son & Signe Oksefjell (eds.), *Corpora and cross-linguistic research: Theory,
method, and case studies* (Language and Computers: Studies in Practical Lin-
guistics 24), 25–50. Amsterdam: Rodopi.

Thunes, Martha. 2011. *Complexity in translation: An english-norwegian study of
two text types.* University of Bergen dissertation. https://bora.uib.no/handle/
1956/5179.

Toolan, Michael. 2001. *Narrative: A critical linguistic introduction* (The INTER-
FACE Series). London: Routledge. Second edition.

Trujillo, Arturo. 1999. *Translation engines: Techniques for machine translation.*
Berlin: Springer.

Tucunduva, Camila de Andrade. 2007. *Translating completeness: A corpus-based
approach.* Florianópolis: Federal University of Santa Catarina MA thesis.

Venuti, Lawrence (ed.). 2000. *The translation studies reader.* London: Routledge.

Vik, Bjørg. 1979. *En håndfull lengsel.* Oslo: J.W. Cappelens Forlag.

Vik, Bjørg. 1979a. *Out of season and other stories.* London: Sinclair Browne. Trans-
lated by David McDuff and Patrick Browne.

Vinay, Jean-Paul & Jean Darbelnet. 1995. *Comparative stylistics of French and English: A methodology for translation* (Benjamins Translation Library 11). Amsterdam: Benjamins. Translated and edited by Juan C. Sager and M.-J. Hamel.