3 Analytical framework

3.1 Overview

This chapter is divided into two main parts. In the first part, 3.2 with subsections, we start by presenting, in informal terms, the information-theoretic concepts of computability and complexity, as well as certain related notions. Next, the relevance of complexity theory for studies of natural language is discussed before various approaches to the notion of 'linguistic complexity' are presented. On that background we introduce our own framework for describing translational complexity, and, finally, discuss the notion of 'computability' in relation to translation.

In the second part, 3.3 with subsections, the correspondence type hierarchy is presented in detail. Each correspondence type is described in terms of (i) the linguistic characteristics of the relation between source and target string, (ii) the amounts and types of information needed to translate, and (iii) the processing effort required by the translation task.

3.2 Computability and complexity

As mentioned in 1.1 and 1.2, the notions of 'computability' and 'complexity' are central concerns in the present investigation. So far in our discussion, 'computable' has been understood as 'solvable by a specifiable procedure' (cf. 2.3.2), and with respect to 'complexity', it has been introduced in the sense of 'translational complexity' and linked with the translator's need for information.[1]

Information science provides the formal tool of computational complexity theory for the purpose of measuring the inherent complexity of computable tasks. For

---

[1] Cf. 1.1, and 1.3–1.3.2.

several reasons this tool cannot be applied to our investigation, but we want to give an informal description of the information-theoretic notions of ‘computability’ and ‘complexity’ in order to reach an intuitive understanding of the concepts. This is meant to throw some light on the motivation behind our concept of translational complexity.

3.2.1 An informal look at the information-theoretic concepts

‘Computability’ is a property of tasks: if a certain task can be solved by a specifiable procedure, then it is a computable task. This means it is possible to write a procedure leading step by step from an initial state to a final state, and in the final state the task is solved. This kind of procedure is called an *algorithm*. Thus, an ‘algorithm’ is a well-defined sequence of steps which always gives a result, i.e. the final state of a computation.[2] Likewise, a ‘computation’ is a step-by-step procedure solving a certain task according to the specifications of an algorithm.

In the context of computational complexity theory ‘complexity’ is a mathematical property which concerns the amount of time and space needed to solve a computable task. Thus, computability is a prerequisite for complexity measurements. We assume that any computation requires certain resources used by the computing device (be it a computer or the human brain), and these resources are processing time and memory space. Computational complexity is a measure of the rate at which a specific computation consumes time and space (van de Koot 1995: 41). In the following we will refer to computable tasks as *problems*, i.e. problems to be solved.

Barton et al. (1987: 7) point out the difference between problem complexity and algorithm complexity. These aspects are in principle independent of each other: it is normally possible to write different algorithms for solving the same problem, and there are cases where the same algorithm may be used to solve different problems. Different algorithms written for the same problem may be more or less efficient than each other, and for this reason the complexity of problems is not measured with reference to specific algorithms.

---

[2] Cf. van de Koot (1995: 40–41).

The task of complexity analysis is to study the structure of a problem, i.e. what Barton et al. (1987: 4) refer to as the *information processing structure* of a problem. This structure determines how the problem can be solved in the most efficient way, i.e. how to specify an algorithm that does not consume larger amounts of time and space than necessary. For instance, given the task of looking up a certain word in a dictionary, this can be done by an inefficient algorithm where the search starts from the beginning of the dictionary, and checks every entry word until the given search word is found. A more efficient algorithm would exploit the fact that a dictionary is an alphabetically sorted list. This could be done by splitting the list into two equal parts, and then checking the beginning and end of each part in order to find out whether the search word is contained in the first or the second part. The search then continues in the relevant half of the dictionary, and the algorithm repeats the splitting into halves until the search word is found. This algorithm is called *binary search*, whereas the former algorithm can be described as search by brute force, or as *exhaustive search*. For obvious reasons a binary search will be more efficient than the exhaustive method unless the search word happens to be located in the beginning of the dictionary. An exhaustive search would be necessary only if the search space is an unstructured list.

Computational complexity theory works not only by analysing the complexity of specific problems, but also by comparing new problems to problems for which the complexity is known. Such comparisons group problems into so-called *complexity classes*, which are classification schemes based on measurements of problem complexity.[3] In this context complexity theory makes a distinction between, on the one hand, problems at type level and, on the other hand, instances of given problems.[4] Such instances can be seen as specific computations where the problem is to be solved for a given input. The amounts of time and space needed to solve a problem are correlated with its size, i.e. the length of the input to a computation. In general, the longer the input the greater the need for time and space. In complexity theory

---

[3] For more information on this, see chapters 1 and 2 in Barton et al. (1987).

[4] See e.g. van de Koot's description of the technique for comparing problems with respect to complexity: "Take a problem of known complexity… Then construct an efficient mapping from instances of the problem of known complexity to instances of the new problem…" (1995: 45).

such correlations can be expressed by mathematical functions relating the size of the problem to the required amounts of time and space, and for the purpose of sorting problems into complexity classes, it is particularly important to know the *order of growth* of these functions (van de Koot 1995: 41). This refers to the rate at which the consumption of time and space increases when there is a growth in the size of instances of specific problems, and this rate serves to group problems into complexity classes. Different classes will have different rates, and classes of harder problems will have higher growth rates than classes of less hard problems.

The sorting of problems into complexity classes provides a precise measure of how hard it can be to solve a problem in typical instances as well as in so-called worst cases. Given a certain complexity class, no algorithm can do better than a certain level of performance. Problem complexity thus delimits the efficiency of optimal algorithms: it is impossible to write algorithms which take less time in the worst case than the amount of processing time required by the inherent complexity of the problem (Barton et al. 1987: 8).

A few main classes of complexity could be mentioned. First, there is a class of problems where the increase in the consumption of processing resources is proportional to an increase in the problem size. That is, if the input grows in size by some integer *n*, then the growth rate of the processing effort is also *n*. In mathematical terms such problems are solvable in linear time, and represent the least hard problems. Second, there is a class of problems where the growth rate can be described as $c \cdot n^k$, where *c* and *k* are constants, and *n* is the size of the problem instance (van de Koot 1995: 41). If the input grows in size by *n*, then the increase in processing time is proportional to $c \cdot n^k$. Such problems are solvable in polynomial time, and within this complexity class there is a considerable degree of variation: cases where the constants *c* and *k* have low values are much less demanding to solve than cases where their values are high, and a sharp increase in the value of *n* will also give a high growth rate. Third, there is a class of harder problems where the growth rate is expressed as $c \cdot k^n$ (van de Koot 1995: 41), so that if the input grows in size by *n*, then the increase in processing time is proportional to $c \cdot k^n$. These problems are solvable in exponential time, which means that only the smallest increase in the problem size causes a great increase in the consumption of processing resources.

The mentioned classes are only a few main categories; several other classes and subclasses have been identified in complexity theory. An important result of this classification is the division between *tractable* and *intractable* problems. We will here not go into the mathematical properties behind this distinction (see Barton et al. 1987: 8–10), but only roughly indicate that tractable problems are solvable within polynomial time, whereas the intractable ones are "problems for which only exponential solution algorithms are known" (1987: 9). In practice, this means that tractable problems are solvable in reasonable time on an ordinary computer, while intractable problems are not.

Closely related to the distinction between tractable and intractable problems is the notion of an 'efficient algorithm'. Algorithms working within the limits of polynomial time are regarded as *efficient*, while algorithms exceeding this upper limit on processing time are not. Keeping in mind the fact that algorithm complexity is independent of problem complexity, it is appropriate to mention the following point made by Barton et al. (1987: 10): "… if a problem is efficiently solvable at all, it will in general be solvable by a polynomial algorithm of low degree."

3.2.2 The relevance of complexity theory for natural language

Human language processing may be seen as instances of computation. In language comprehension the human brain processes the input in order to construct an interpretation of it, and speech production involves producing an output for the purpose of expressing some intended meaning. These are types of computation where the brain uses time and memory to process information drawn from linguistic knowledge as well as from knowledge of the world. Such processes fall within the study of psycholinguistics, and shall not be dealt with here.

Viewing human language processing as computation makes it natural to apply the tools of computational complexity theory to natural languages, and this is clearly a way of gaining insight into the challenges mastered by the human language ability. A different avenue of research is to study frameworks for language descriptions, i.e.

grammar formalisms, in terms of computational complexity. Van de Koot (1995: 38) mentions several publications dealing with the application of complexity theory to natural language and linguistic theory.

With respect to complexity analysis of natural languages, we may briefly present a few contributions found in the literature. Because it is difficult to investigate the algorithms used by the human brain for language processing, little is known about the very processes.[5] Hence, Barton et al. (1987: 2) point out that in order to study human language processing, computational complexity analysis is a most appropriate tool since it is independent of solution algorithms as well as of computing devices. In the case of human language processing we have access to its input and output, i.e. natural language, but we do not have direct access to the inside workings of the processes themselves. That is, however, no problem for complexity analysis since it pertains to problem structure (cf. 3.2.1) and may be applied to available linguistic data. Barton et al. (1987: 4) hold the view that "…there is every reason to believe that natural language has an intricate computational structure that is not reflected in combinatorial search methods."[6] In other words, they believe that if the outcome of a natural language problem is intractability, then it is likely that some of the structure of the problem has not been detected.

Van de Koot (1995: 39) observes the following with respect to the complexity of natural languages: "The picture that emerges … is that natural language computations are computationally intractable … But … they have the useful property of being on the verge of tractability: their solutions are hard to find but easy to check once found." Van de Koot (1995: 39) makes an interesting point about the usefulness of applying complexity analysis to natural language computations: studying the computational complexity of language problems such as comprehension makes it possible to "relate language computations to other computations whose structure we understand and provide a design target for language algorithms (i.e. for algorithmic characterizations of language computations)." This means that complexity analysis of natural

[5] Cf. the discussion of process-oriented translation studies in 1.4.1.3.

[6] Combinatorial search is search by brute force, i.e. to try all possible combinations; cf. 3.2.1.

language problems may facilitate the design of algorithms for natural language processing within the field of human language technology.

The application of complexity analysis to grammar formalisms also deserves to be mentioned. It is regarded as necessary requirements of grammar formalisms that they, on the one hand, have the sufficient means for expressing all possible structures in any language, and that they, on the other hand, rule out the description of structures that do not belong to any possible language. These requirements are concerned with what is referred to as the *generative capacity* of grammar formalisms, or their generative power. A consequence of the view of Barton et al. (1987: 4) that natural language has a computational structure that does not require combinatorial search methods, is that a grammar formalism should not be able to express linguistic structures for which the comprehension task would require exponential solution algorithms. If that is the case, then the formalism is too "powerful"; it will generate more than the structures found in natural languages. According to Barton et al. (1987: 4), complexity analysis can be used, then, to identify the parts of a grammar formalism that allow the generation of linguistic structures more complex than natural ones. In other words, complexity theory can be used to weed out over-generation in grammar formalisms, and this is a very useful tool for the study, and development, of linguistic frameworks for language description.

3.2.3 Linguistic complexity

Having discussed the information-theoretic notion of 'complexity', we now move on to complexity in the context of language. This is not a computational concept; it may (although not exclusively) be related to learning rather than to processing, and it may be related to subjective experience, whereas computational complexity is an objective property of a computable task.

In the structures of natural languages we may intuitively perceive different degrees of complexity, for instance in connection with language learning. A language with a rich inflectional system may be experienced as very complex by a foreign language learner if his or her mother tongue is a language with little morphology. E.g., a Norwegian learner of German will normally find it challenging to acquire command of four different types of grammatical case in German, since Norwegian exhibits no grammatical case distinctions apart from an opposition between nominative and accusative in pronouns. And for a Norwegian the Finnish case system, needless to say, seems overwhelmingly complex with its fourteen different kinds of grammatical case.

An interesting challenge for linguists, then, is to find a principled way of describing linguistic complexity. This opens up for questions such as what kinds of phenomena, or which aspects of language, are involved in linguistic complexity, and how can a given structure be described as more, or less, complex than other structures it may be compared with? In 3.2.2 we have presented computational complexity theory as a tool suitable for studying complexity in natural language, but we have not mentioned that such studies cannot be done without certain prerequisites. Since complexity analysis is, basically, to analyse the structure of a problem for the purpose of finding a solution to it, it is necessary to transform language into a computational problem before it can become the object of complexity analysis. Firstly, this requires a computationally implementable formalism for language description, and, secondly, a grammar written in that formalism for the relevant language. Thirdly, it is necessary to reduce the study of structures in that language to the task of deciding whether the structures belong to the language described by the given grammar. This is commonly called a *recognition* problem,[7] and it presupposes a conception of language as the set of strings, or expressions, generated by a certain grammar. Complexity analysis is then applied to the recognition problem, and the result of the analysis serves eventually as an estimate of the complexity of the given language structures. In other words, using complexity analysis to investigate linguistic complexity involves the construction of search tasks performed on the basis of formal grammars.

As a consequence, there are many language researchers who want to study linguistic complexity, but do not apply computational complexity analysis. The field may not be seen as relevant, or the necessary prerequisites may be lacking. But irrespective of method, such studies anyway need a clear understanding of

[7] See e.g. van de Koot (1995: 46ff).

complexity, and several researchers have made contributions in this respect. One example is Dahl (2004), whose approach to linguistic complexity is based on information theory, in particular the principles of signal transmission. His project is to study, from a diachronic perspective, how the complexity of language systems evolves and is maintained. In the context of language, Dahl describes the notion of 'complexity' as "[not] synonymous with "difficulty" but as an objective property of a system — a measure of the amount of information needed to describe or reconstruct it" (2004: 2). With respect to the information-theoretic notion of 'complexity', he says informally that "the complexity of an object would … be measured by the length of the shortest possible specification or description of it" (2004: 21).

Dahl points out that available background information may have consequences for how long the shortest possible description needs to be, and this calls for a distinction between absolute and relative complexity (2004: 25–26). The complexity of an entity *relative* to a certain amount of available information is measured by the length of the specification needed, in addition to the available information, to describe the entity. *Absolute* complexity, on the other hand, pertains to the total amount of information needed to describe the entity.

In his approach to complexity in languages, Dahl describes a language as consisting of resources and regulations: resources are the building blocks of linguistic expressions, while regulations determine how the resources are used correctly (2004: 40–42). He then argues that measuring the complexity of the resources of a language is distinct from measuring the complexity of its regulations. The former pertains to the size of the inventories that are included in the resources, whereas the latter relates to the complexity of the expressions of the language, and is understood by Dahl as *system complexity* (2004: 42–43), which is the central concern of his study of complexity in languages.

Miestamo (2006) presents other approaches to studies of linguistic complexity, and his context is the study of language typology, and in particular studies where languages are compared with respect to the complexity of specific domains of grammatical functions, such as aspect, tense, negation, or definiteness. He, too, applies an opposition between absolute and relative complexity in natural language, although his explanation of relative complexity is somewhat different from that of Dahl (2004):

> "The absolute (or theory-oriented) point of view looks at complexity in terms of the number of parts in a system, or in information-theoretic terms (Shannon 1948) as the length of the description a phenomenon requires (cf. Dahl 2004). The relative (or user-oriented) point of view pays attention to the users of language and defines as complex what makes processing, acquisition or learning more difficult." (Miestamo 2006: 346)

Miestamo (2006: 348–349) argues that to be able to study linguistic complexity in general terms, it is not sufficient to investigate complexity in relation to one group of language users and not another. The reason is that it appears to be arbitrary which group(s) of language users experience(s) a certain linguistic property as difficult, or complex (cf. the above example of grammatical case). Hence, it is problematic to describe linguistic complexity relative to language users, and Miestamo's conclusion is that the absolute, or theory-oriented, approach is more fruitful in order to achieve precise criteria for the description of linguistic complexity.

Like Dahl (2004), Miestamo applies an informal interpretation of the information-theoretic notion of 'complexity' when he defines a complex phenomenon as "something requiring a longer description than a less complex phenomenon" (2006: 349). He argues that this definition provides an objective criterion for cross-linguistic comparisons with respect to specific grammatical properties, and in Miestamo (2006) it is applied to a typological study of the functional domain of negation. More specifically, it is used to study the relationship between grammatically expressed functions and their formal encoding, and Miestamo observes that "a language where more grammaticalized distinctions are made in a given functional domain, requires a longer description for that functional domain than a language where less distinctions are made" (2006: 349). This is the standard of measurement for cross-linguistic comparison, and it is a quite separate topic whether it is difficult to comprehend or acquire the linguistic distinctions expressed in a given domain. The latter illustrates the relative point of view in relation to linguistic complexity.

3.2.4 Translational complexity

Our approach to complexity in translation is an attempt at creating a fairly precise frame of reference for the characterisation of translational correspondences, although it is not an aim to find a mathematically exact description of complexity in relations of translation. We cannot apply the tools of computational complexity analysis, as the prerequisites needed for that (cf. 3.2.3) are not available in our project. Like Dahl (2004) and Miestamo (2006), we adhere to an absolute point of view in our description of complexity: translational complexity is analysed in terms of a quantifiable, objective commodity (information), and independently of the competence of the translator who has produced a specific translation. Our approach differs somewhat from those adopted, respectively, by Dahl (2004) and Miestamo (2006) in the sense that they measure linguistic complexity as *length of description*, whereas our analysis of translational complexity is based on *amounts and types of information* needed in translation, and in several respects our investigation rather resembles the techniques of computational complexity analysis. However, since we are not in a position to quantify information in exact terms, we have to rely on more intuitive notions.

Firstly, complexity analysis applies to computable tasks, described as problems (cf. 3.2.1), and in a similar way we want to characterise the degree of complexity in translational correspondences by viewing them as *translation tasks*, i.e. the task of producing a particular target expression on the basis of the information encoded in the given source expression together with other information sources. This could more precisely be seen as describing the complexity of a specific solution to a translation task.[8] Then, the notion of a 'search task' becomes a common denominator between computational problems and translation tasks. An algorithm solving a problem carries out a search task, i.e. the search for the solution, or the final state of the computation. Likewise, a translation task can be regarded as a search task: the search for the

[8] This is a narrower notion of 'translation task' than the one presented in 2.4.2.2. The concept is further discussed in 3.3.1.1.

information needed to interpret the source text correctly, and for the information needed to produce the target expression.

Next, complexity analysis uncovers the structure of a problem for the purpose of finding a solution to it, and this analysis is independent of possible algorithms that may solve the problem. As a parallel to this, we study the linguistic relation between source and target expression in translational correspondences irrespective of how the translation has been produced. Given our view of translation as a search task, this is an analysis of the *structure* of the search task. In this context we may regard the source expression as the initial state of the translation task, and the target expression as its final state. Since we do not have direct access to information about the translation process (cf. 1.4.1 and 1.4.1.3), what we can do, in order to measure the complexity of a specific translation task, is to study the relation between the initial and the final state by analysing how source and target expression correspond (or not) with respect to linguistic properties. On the basis of syntactic and semantic correspondence relations between source and target string (divergences included), the structure of a translation task can be analysed in terms of the types and amounts of information needed to solve it. We emphasise that this approach to translational complexity is no attempt to describe aspects of the translation process or any possible algorithm producing the translation. Rather, it aims at capturing the necessary requirements for solving the translation task.

Further, there is a parallel between our analysis and complexity theory with respect to the measuring of complexity. In complexity theory this is measured in terms of the consumption of resources, i.e. processing time and memory space, needed by a computation (cf. 3.2.1). These factors reflect the *processing effort* required by the computation. In our analysis translational complexity is measured in terms of the need for information in a translation task, and the processing of this information is a kind of computation that will consume time and space. As already made clear, it is, however, not possible in our study to calculate numerically the precise amounts of processing resources needed for specific translation tasks, but we do distinguish between classes of translational correspondences according to certain assumptions regarding the consumption of resources, or processing effort. These assumptions may be said to concern the *weight* of the search task and pertain to two closely related topics: (i) the extent to which the various types of information needed to carry out a translation task can be represented in a finite way, and (ii) the amount of effort required in order to access and process them. Naturally, the processing of easily accessible information requires a smaller effort than the processing of less easily accessible information. E.g., information that can be looked up as easily as information presented in a table, is directly accessible, while the accessing of information that must be derived through linguistic analysis of an expression, or through inferencing, involves a greater effort. These topics will be revisited in the presentation of the correspondence type hierarchy later in this chapter.[9]

Finally, our sorting of types of translational correspondences resembles the way in which complexity theory groups computable tasks into complexity classes. As explained in 3.2.1, complexity classes are distinguished on the basis of the rate at which the consumption of resources grows in proportion to an increase in the size of the computable task. The growth rate indicates how hard it is to solve the problems of a given class in the best, typical, and worst cases, respectively. The presentation of the correspondence type hierarchy will show that the four types are distinguished in terms of a lower bound on how easy it can be to produce the target text in a translation task, i.e. which types of information, and how large amounts of them, the translator must *at least* have access to within a certain class of correspondences. The four types are also distinguished in terms of an upper bound: each correspondence type is associated with a certain set of sufficient information types, so that translation tasks requiring information types in addition to the set associated with a given correspondence type, are of a more complex type. In the present approach differences in the degree of translational complexity cannot be measured *within* the different correspondence types. Still, it is our hope that the distribution of the different correspondence types in a body of parallel texts may reflect the degree of translational complexity in an interesting way.

---

[9] These two topics have previously been mentioned in 1.2 and in 2.4.2. The issue of finiteness concerns the prestructured domain of linguistic information which defines the limit of predictability in the translational relation; cf. 2.3.2. The issue of processing effort has so far been mentioned in general terms only in connection with the information typology presented in 2.4.2 with subsections.

3.2.5 Computability in relation to translation

The notion of 'computability' is defined in 3.2.1 as a property of tasks: a task that can be solved by a specifiable procedure is computable. By *computing a translation* we understand creating a target text by means of software for automatic translation and without the aid of a human translator. Previously, in 2.3.2, we have drawn a principled limit on computability in the translational relation between two languages: this relation is computable insofar as it is *linguistically predictable*, i.e. to the extent that the translation of a given source text can be predicted, as one possibility, by means of the information encoded linguistically in the original, together with pre-structured information about the source and target language and their interrelations. These interrelations represent the translational relation on the level of system units, which is distinct from the translational relation on the level of linguistic usage (cf. 2.3.1). As explained in 2.3.2, there are instances of translation that cannot be accounted for by pre-structured linguistic information alone; hence we assume that the extension of the translational relation on the level of *langue* is a subpart of the extension of the translational relation on the level of *parole*. This entails the further assumption that there is a part of the translational relation that cannot be automatised, since we regard the *langue* relation as the limit on computability.[10] With reference to automatic translation, we thus understand the *computable* part of the translational relation to be the same as the *linguistically predictable* part of it, and computable translation tasks are solvable within the pre-structured domain of linguistic information, as previously explained in 2.3.2.

In 3.2.2 we have argued for viewing human language processing, of which translation is one kind, as computation. The translational relation on the level of *parole* covers what bilingually competent humans are able to translate, which means that in relation to human translation, we will by 'computable' understand 'translatable'. In the same way as computability is a prerequisite for computational complexity measurements, translatability is a prerequisite for analysing translational complexity. To avoid confusion, we will in the context of translation restrict the

---

[10] I.e., we assume that only the linguistically predictable translations are computable.

notion of 'computability' to the linguistically predictable part of the translational relation, and use the notion of 'translatability' in connection with translation tasks that humans are able to solve. Although we find it appropriate, in the context of complexity analysis, to view human translation as a partly computable task, we will not in general speak of the human activity of translation as *processing* or *computation*, as such analogies do not contribute to keeping a clear distinction between human and automatic translation.

The notion of 'translatability' can hardly be mentioned without evoking the question whether there is a limit to translatability — whether any translation task can be solved by sufficiently competent human translators. There may be cultural differences, as well as differences between source and target language systems, which may force the translator to paraphrase a given source expression to the extent that the target expression appears as a rewriting rather than a translation. It is then a question of definition how much a target text may diverge from the source while still functioning as a translation.[11] The discussion of the translatability issue falls outside the present project, but has received considerable attention within translation studies. Our focus is on delimiting the computable within the translatable.

3.3 Translational correspondence types

Our empirical investigation, to be described in chapter 4, is a classification of translationally corresponding strings into four different types, introduced in 1.3 with subsections. In simplified terms, type 1 correspondences are cases of a full linguistic match, structurally as well as semantically, between source and target string; type 2 correspondences allow minor mismatches on the structural level, but none on the semantic; in type 3 correspondences there can be major structural divergences while there is still a semantic match, and in type 4 correspondences there are semantic as well as structural mismatches between source and target string. Instances of each correspondence type are identified through syntactic and semantic criteria, and the types are related to each other in a hierarchy, reflecting an increase, from type 1 to 4,

[11] Cf. the prototypical view of 'translation' in Halverson (1998), discussed in 1.4.1.

in the amount of information necessary to produce the target expression. This need for information is in turn an indication of the degree of translational complexity in the translation task. Given the assumptions of our analytical framework, correspondence types 1, 2, and 3 fall within the limit of computability in the translational relation, whereas type 4 correspondences are not computable as they involve information sources not included in the pre-structured domain of linguistic information. In 3.3.2–5 with subsections, the correspondence type hierarchy is presented in detail.

3.3.1 General aspects of the classification of translational correspondences

Certain topics involved in the classification of translational correspondences are relevant to the whole set of correspondence types. To avoid repetition in the presentations of each type, these topics are discussed in 3.3.1.1–4.

3.3.1.1 The notion 'translation task'

The notion of 'translation task' is important in our approach to translational complexity. The concept is used in chapters 1 and 2 in a general sense which covers the task of translating anything from a single lexical item, or a sentence, to an entire document, such as a handbook or a novel.[12]

Section 3.2.4 introduces a more precise sense, where a 'translation task' is understood as the task of producing a particular target expression by means of various sources of translator's information together with the information encoded in the source expression, given its relevant interpretation. This can be seen as the specific task of translating a textual occurrence of an expression *a* of a certain source language (L_1) into expression *b* of a given target language (L_2), i.e.: $a_{L1} \rightarrow b_{L2}$

It should be noted that we have decided to keep the subtask of source text disambiguation apart from the translation task. The reason is that our analysis of complexity pertains to translation tasks only, and hence we do not consider the problem of source text disambiguation. By the *relevant interpretation* of a_{L1} we here mean the interpretation which lies behind the chosen translation b_{L2}. How the

[12] The general notion is commented on in 2.4.2.2.

translator has identified the relevant interpretation falls outside the scope of our analysis.

There is also a somewhat different, and extensionally wider, sense of 'translation task', namely the task of translating a textual occurrence of a certain source expression a_{L1}, given its relevant interpretation, into a specific target language L_2. In such general translation tasks the source expression a_{L1} corresponds with a set of possible target expressions (T_{L2}), i.e.: $a_{L1} \rightarrow T_{L2}$. Thus, whereas translation tasks of the first kind involve correspondences between specific, single expressions of L_1 and L_2, translation tasks of the second, more general, kind involve correspondences between specific expressions of L_1 and sets of translations in L_2, from which the translator makes a motivated choice.

Notably, what we have aimed at in the analysis of translationally corresponding string pairs is to measure the complexity in a collection of concrete translation tasks (i.e. string pairs) where the chosen target expression is only one of a set of possible translations in L_2. Thus, the complexity measurement applies to specific translation tasks $a_{L1} \rightarrow b_{L2}$, and the analysis of each string pair is an attempt to describe the complexity of the selected task *solution* in relation to the source expression a_{L1}, given its relevant interpretation. We do not consider the complexity of the translation task that is not solved yet; that would amount to analysing the complexity of the general translation task ($a_{L1} \rightarrow T_{L2}$), which has a set of possible solutions.

In the presentation of the correspondence type hierarchy each type is related to the specific notion of translation task. The purpose of describing the hierarchy in terms of translation tasks is to explicate the information requirements of each type, i.e. the types and amounts of information needed in order to produce the chosen solutions to specific translation tasks. Still, it should be noted that when we refer to the task of producing a particular translation from a given source text, the notion of a 'task' cannot be related directly to the translator's situation. In many cases the translator might have chosen less complex (i.e. literal) solutions, which means that the task of translating the given source expression in general may be simpler than the solution actually chosen. Hence, what we aim to describe is the complexity and information requirements of a specific solution to a translation task. In the correspondence type descriptions we will discuss complexity measurement, or type identification, as part of the task. By this we do not mean the identification of the simplest possible type of solution to a given task, but the identification of the complexity type (1, 2, 3, or 4) of the solution that has been chosen by a translator.

Although the orientation of the present study is different from that of Toury (1995), there is an interesting parallel between our approach to translational correspondences and his notion of 'coupled pairs'.[13] Toury (1995: 77) defines 'coupled pairs' as correspondences between specific translation problems in the source text (i.e. tasks to be solved), and their solutions in the target texts. In his view, such coupled pairs should be the starting point for the description of translational phenomena, and he emphasises that in coupled pairs, source problems and target solutions "should be conceived of as determining each other in a mutual way" (1995: 77).

As stated in 3.2.5, our investigation aims at delimiting the computable part of the translational relation within the domain of the translatable. Information about source and target language and their interrelations defines the linguistically predictable, or computable, set of correspondences between SL and TL. When we analyse the degree of translational complexity in selected parallel texts, we describe how translation tasks are solved by certain translators, and the classification of translational correspondences is thus meant to reflect the complexity in the task of generating automatically the translations that some humans have produced.

3.3.1.2 Criteria for distinguishing and describing correspondence types

In the present approach a set of three criteria is used to distinguish between the four types of translational correspondences. The first criterion pertains to the linguistic characteristics of the relation between source and target string, characteristics which show the degree to which there exist implications between relations of equivalence between source and target string. The second criterion concerns the amounts and

[13] As previously observed in 1.4.1.1, Toury describes his study as "an attempt to gradually reconstruct both translation decisions and the constrains under which they were made" (1995: 88), and this is his motivation for identifying units of comparative analysis.

types of information needed to produce the translation, and may be conceived of as the *structure* of the search task involved in translation. The third criterion deals with the processing effort required by the translation task, which may be seen as the *weight* of the search task.[14] As explained in 3.3.1.1, each correspondence type is to be described in terms of the notion of translation task, which will be decomposed into the three subtasks of source text interpretation (or analysis), complexity measurement, and target text generation. The subtask of complexity measurement will be referred to as *type identification* in the presentations of the four correspondence types.

Firstly, translational correspondences are classified in terms of the linguistic properties of the relation between source and target strings: these properties are the criteria through which tokens of each correspondence type may be identified. If there is some degree of structural similarity in a given language pair, then there will be a certain set of linguistic structures in the source language sharing properties with translationally corresponding structures in the target language. Information about such correspondence relations is included in the general information about interrelations between source and target language (cf. 2.4.2.2). I.e., we assume that information about how constructions in the two languages are translationally related (or unrelated) to each other is information available prior to translation. In cases exhibiting a high degree of structural relatedness between the source and target expression, the translation task is easy to solve, while it is harder in cases where original and translation are structurally unrelated.

The presentation of the correspondence type hierarchy will show that in cases where similar structures of respectively SL and TL are translationally matched, there will exist relations of equivalence between source and target string, and, also, implications between such equivalence relations.[15] These relations of equivalence concern different linguistic levels: syntax, semantics, and pragmatics. The discussion of the correspondence types will illustrate that in cases where source-target equivalence with respect to syntax implies equivalence also with respect to semantics and pragmatics the degree of translational complexity is low, and that as translational

---

[14] Cf. 3.2.5, where we have previously commented on the structure and weight of translation tasks.
[15] Dyvik (1999: 229–230) describes translational complexity in terms of such implications.

complexity increases, such implications exist to a lesser degree. The domains within which such implications hold are assumed to be delimited by information about the translational relationship between source and target language. In this context 'equivalence' should not be taken as identical to the notion of 'translational equivalence' discussed in 1.4.1.1, but rather be understood as 'linguistic matching relations'. We have nevertheless chosen the expression *equivalence* since we regard it as more precise than *match*, and in this context it may be understood as equivalence between original and translation with respect to specific linguistic properties.

Secondly, each correspondence type will be characterised with respect to the amounts and kinds of task-specific information required to translate source language strings. According to the discussion of translational complexity in 3.2.4, this may be interpreted as an analysis of the structure of the search task involved in translation. The search task is twofold: there is, first, the search for the information needed to interpret the source text correctly, and, second, the search for the information required for producing the target expression. Solving the first subtask, interpretation, involves using the information encoded in the source string together with information about the source language system. From the perspective of computing the translation, there is also a need for an intermediate subtask of complexity measurement, i.e. diagnosing the degree of complexity in the translation task. This requires information about the linguistic structure of the source string and information about the interrelations between the source and target language systems.[16] The final subtask, generation of the target string, requires information about the interrelations between source and target language, as well as information about the target language in isolation.

With respect to these subtasks, we shall discuss how the need for information is correlated with the degree of translational complexity in the different correspondence types. The decomposition of the translation task into analysis, complexity measurement, and generation should not be taken as assumptions concerning how a translation task is solved by a human translator; rather, it is a description of an idealised,

---

[16] This topic is discussed further in 3.3.1.4.

minimal procedure on which possible translation algorithms may be based, i.e. a description of the information processing structure (cf. 3.2.1) of the task of computing a translation.

Thirdly, each correspondence type is characterised with respect to what is described in 3.2.4 as the weight of the translation task, i.e. the amount of required processing effort. For each correspondence type the necessary information sources are for this purpose viewed in relation to two topics: the extent to which they can be represented in a finite way, and the amount of effort required in order to access and process them. The decomposition of the translation task into three subtasks is relevant also for these topics as the amount of required effort varies not only among the types of translational correspondences, but, as we shall see, also among the subtasks.[17]

3.3.1.3 The notion 'necessary information'

The string pairs we have analysed are produced by human translation, and hence they represent translation tasks solvable by bilingually competent language users (insofar as each source string does have a corresponding target string). A subset of these string pairs represent computable tasks, which can be solved, given certain assumptions, by pre-structured linguistic information sources alone. In the classification of string pairs into correspondence types it is an aim to identify the information sources that are *at least* necessary in order to compute (or, if not computable, to produce "manually"), each target string.[18] With respect to the subset of computable translation tasks, we will argue that in some cases it is not necessary, in order to generate the target text, to analyse the source text further than to the level of syntax (types 1 and 2), whereas in other cases a semantic analysis of the source text is also required (type 3).

Again, this must not be interpreted as a way of conceptualising the translation process. It is not plausible that a human translator, after having read a text string in the source language will consider only its syntactic structure, and disregard its semantic content as well as accompanying contextual information, because he or she

[17] Cf. 3.3.1.4, as well as the discussions of each correspondence type.
[18] Cf. the remarks in 3.2.4 on how the correspondence types are distinguished from each other.

is aware that a target string with a structure identical to that of the source string is an appropriate translation. Rather, a competent translator must continually pay attention to the meaning and context of the source text, and in cases where he or she chooses a word-by-word translation, that is done especially because it seems appropriate after having considered the meaning and context.

But from the perspective of a system for automatic translation, it is a formidable task to process the various types of information associated with even a very short text. To analyse the syntactic structure of a limited source string is, on the other hand, a computationally tractable problem, and hence a good starting point for identifying the degree of complexity of the given translation task. If the system is able to decide that the target language offers syntactic structures matching those found in the source string and with corresponding compositional semantic properties, then we assume that an efficient strategy for automatic translation is to refrain from analysing the meaning and context and simply proceed to identifying the corresponding target words and generating the translation directly from the source string.[19] Thus, our attempt at identifying *the necessary information sources* for translation in relation to each correspondence type is a way of describing how the complexity of chosen translation task solutions is determined by how much and what kinds of information that must *at least* be available in order to produce them.

3.3.1.4 The need for general information sources

As discussed in 2.4.2.2, we assume for each correspondence type that certain general information sources are available prior to the translation activity, i.e. information about source and target language and their interrelations, and various kinds of extra-linguistic background information. These sources exist independently of specific translation tasks, but constitute an important part of the total amount of information needed to solve a given task. As already indicated in 3.3.1.3, the different correspondence types vary with respect to how much of the given information sources are required. Granted that types 1–3 represent translation tasks solvable within the pre-

[19] This is one of the central design principles of the PONS system (Dyvik 1990, 1995). Cf. the description in 1.3.2 of the different modes of translation in that system.

structured domain of linguistic information, it is only in type 4 correspondences that other information sources are needed to produce the target text. Moreover, with regard to the types 1–3, we shall see that syntactic and morphological information is sufficient in types 1 and 2, while in type 3 semantic information is also required in order to compute translations.

As discussed in 2.3.2, linguistic and extra-linguistic information sources differ in the sense that the former represents a limited domain, whereas information about the world is an open-ended domain. Thus, given the scope of general language, it is theoretically possible to represent information about source and target language and their interrelations in a finite way, while there is no principle available in order to determine which pieces of world information to include in information modules for automatic translation systems. In cases where translation requires the processing of given, general *world* information, we assume that, in general, this is not a problem that the computer can solve: the information is not available in the pre-structured domain of linguistic information, and hence not accessible. It is only within artificially delimited domains that world information can be made accessible in finite ways.[20] For the human translator, on the other hand, it is hardly an effort to make use of general, extra-linguistic background knowledge.

Given, general *linguistic* information sources are needed in all translation tasks. The need for general linguistic information can be discussed in relation to the division of the translation task into three subtasks: analysis, complexity measurement (or type identification), and generation (cf. 3.3.1.2). Insofar as each of these subtasks requires the processing of given, general *linguistic* information sources, we assume that this is a challenge that the computer can handle, since the information is finite and directly accessible as it is given prior to translation. The amount of processing effort will be determined by the amount of general linguistic information that is needed and the complexity of the task of retrieving it, and on this point there are differences between the subtasks.

---

[20] Cf. the discussion of restricted semantic domains in 1.4.2.3.

Irrespective of the type of translational correspondence, the subtask of analysis requires syntactic parsing of the source string, and the parsing problem is solved by using information contained in the representations of the source language lexicon and grammar. The amount of information that must be accessed is correlated with the length and linguistic complexity of the source string. The first step in parsing is to recognise word forms, and we assume that the information structure representing the SL lexicon is organised by base forms, so that for each inflected word morphological analysis is necessary to identify the lexeme it belongs to. Thus, recognising uninflected word forms requires smaller computational resources (i.e. a smaller number of calculations) than identifying word forms with inflection. Subsequent to word recognition, information about each lexical item can be merged with information about possible syntactic structures of the source language in order to create a parse of the source string, i.e. a representation of its syntactic structure. With respect to the amount of processing effort required by parsing, several researchers have studied the computational complexity of parsing problems, e.g. Barton et al. (1987) and van de Koot (1995). We will not go more deeply into parsing and the topic of its computational complexity, since the complexity of the translation task solution is determined by the relation between source and target string, and not by the complexity of parsing problems, because the parsing task is common to all four types of translational complexity. Thus, the basic amount of parsing needed for all kinds of source strings does not contribute to distinguishing between degrees of translational complexity; it is only parsing tasks associated with certain translational correspondence types, such as the retrieval of semantic information in type 3, which can be seen as contributing specifically to translational complexity.[21]

With reference to the typology of information sources for translation (cf. 2.4.2 with subsections), it is worth noting that while the input to the analysis step is general, given linguistic information (together with the word forms of the source

[21] As regards type 3, cf. 3.3.4.2–3 on this topic. Otherwise, due to the tendency that the high degree of structural relatedness found between source and target strings in types 1 and 2 is more likely to occur in short and structurally simple expressions than in longer and more complex ones, it is the normal case that parsing requires a smaller effort in correspondences of the two least complex types than in more complex cases. This is, however, a contingent aspect of the string pairs, and is in principle independent of the factors that contribute to the degree of translational complexity.

string), the output of the analysis step — i.e. an interpretation of the source string — can be seen as task-specific linguistic information. It is a representation of the information which is linguistically encoded in the source string, and it includes the described situation type as well as information about the linguistic structure of the source string.

The second subtask in the computing of a translation is to measure the complexity of the translation task, and this is done by combining the task-specific linguistic information given in the interpretation of the source string with general information about the interrelations between source and target language systems. These are interrelations between translationally corresponding elements of the lexicons of the two languages as well as between translationally corresponding structures described by rules of the respective grammars of SL and TL.[22] Here it is relevant to explain how information about such interrelations can be made directly accessible to the subtask of complexity measurement.

In the computational perspective it is rational, for a given language pair, to calculate such interrelations once, and store them, so that that information is available, and directly accessible, prior to any translation task. This amounts to computing a comparison of the language descriptions representing respectively SL and TL. We assume that when the interrelations between source and target language are calculated, it is possible to reveal not only between which elements of the two languages there exist translational correspondences, but also to what extent there are relations of equivalence between the corresponding elements, i.e. to determine the linguistic properties which are shared by source and target elements. Once such interrelations between two language systems have been calculated, information about them can be associated with individual elements of the lexicons and with individual rules in the grammars of respectively source and target language.[23] Such information may be seen as describing the translational properties of the individual lexemes and rules in the source language *with respect to the given target language*. Thus, informa-

---

[22] As regards lexicon information, we assume that interrelations between the word inventories of two languages normally apply to lexemes, and not to word forms, as more than one word form may be associated with one lexeme in languages with inflection.
[23] This approach has been tested in the PONS system for automatic translation (Dyvik 1990, 1995).

tion about source-target interrelations is directly accessible when monolingual information about lexical elements and grammatical structures is processed in order to interpret, or analyse, the source text in a given translation task.

This means that when a translation is computed, the subtask of analysis provides the bilingual information needed to diagnose the complexity of the translation task. The underlying principle is that information about how SL and TL are interrelated entails information about translational correspondences between specific linguistic elements in the two languages, so that identifying a particular lexeme or a particular syntactic structure in a source text will provide direct access to information about translationally corresponding elements in the given target language and information about linguistic properties shared by source and target elements. This kind of information is the basis for measuring the complexity of a given translation task, and we assume that the effort involved in accessing and processing such information is comparable to the effort required by the computable task of table lookup.[24] The presentations of each type of translational correspondence will provide further details on how the subtask of complexity measurement (or type identification) is solved.

Finally, with respect to the need for general linguistic information, the subtask of generation requires information retrieved from the representations of the target language lexicon and grammar. The amount of necessary information, as well as required processing effort, will be commented on in connection with each type of translational correspondence.

3.3.2 Type 1 correspondences

In 1.3.1 correspondences of type 1 are described as "word-by-word translations", and they represent the least complex class of translational correspondences. With respect to the language pair English-Norwegian such cases are not very frequent, and the frequency would be higher in language pairs with a greater degree of structural relatedness between SL and TL.[25] The example given in 1.3.1 is here repeated in (1):

---

[24] Intuitively, table lookup demands very small computational resources. It requires no derivations or processing other than reading off the table the information that is available for a given search key.

[25] The proportion of type 1 correspondences within the recorded data is given in table 5.1 in 5.2.1.

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

3.3.2.1 Linguistic characteristics of type 1

Type 1 correspondences are cases where translationally matched structures of respectively source and target language are so similar that there is equivalence between source and target string with respect to the sequence of translationally corresponding surface word forms. For such a string pair to count as a type 1 correspondence, some further requirements need to be fulfilled. Firstly, the strings must be syntactically equivalent, i.e. equivalent with respect to the assignment of syntactic functions (subject, object, etc.) to constituents.[26] Secondly, the syntactic structures have to be compositionally equivalent in the sense of having corresponding properties with respect to compositional semantics: predicates and arguments must be contributed by corresponding constituents. Such compositional equivalence will in the normal case be a consequence of syntactic functional equivalence. Finally, the strings have to be pragmatically equivalent in the sense of being used to perform corresponding pragmatic functions, or speech acts, in the given texts.

These requirements specify to what extent source and target string must exhibit corresponding linguistic properties in order to be classified as a type 1 correspondence. Word-by-word correspondences do not qualify as type 1 unless they also correspond syntactically, semantically, and pragmatically in the way described here. Within a domain of type 1 correspondences delimited by these requirements, there will hence exist relations of implication that can be exploited in the translation process: the fact that there is a type 1 correspondence between source and target string includes the fact that the existence of syntactic equivalence implies semantic equivalence, and that semantic equivalence further implies pragmatic equivalence.[27]

---

[26] There may be differences of detail in the phrase structure trees, as motivated by differences between SL and TL. Further development of this point requires illustration by means of language descriptions implemented in specific grammar formalisms, which we will not do. Anyway, such differences must not violate the requirement of syntactic functional equivalence.

[27] These relations of implication must not be understood as causal relations, but rather as material implications of the form "if *a* is true, then *b* is also true."

As explained in 3.3.1.2, this information is entailed in general, given information about the interrelations between source and target language, and it can be exploited so that the translation task is solved simply by translating word by word and without doing a deep linguistic analysis of the source string.[28]

Given the extent to which linguistic properties are shared between original and translation in type 1 correspondences, in particular the sharing of semantic properties, it follows that type 1 correspondences are included among the linguistically predictable translational correspondences, as described in 2.3.2. That is, a target string corresponding to the source string according to type 1 requirements is a member of the LPT set of the source string.

3.3.2.2 The structure of the translation task in type 1: information sources

Since type 1 correspondences are included among the linguistically predictable translational correspondences, a translation task of type 1 is solvable within the pre-structured domain of linguistic information. It may appear, from a computational point of view, that in type 1 the translation task merely involves replacing the word forms in the source string with the translationally matching word forms of the target string.

However, interpreting the source string is an initial, indispensable subtask, especially since it is required to determine that the given translation task is a type 1 case. The point was made in 3.3.1.3 that because a deep linguistic analysis of text is computationally resource-intensive, we assume that a rational strategy for computing a translation is to determine the amount of work required to generate the target text, i.e. to measure, or diagnose, the degree of translational complexity. In order to identify a translation task as a type 1 case, it is necessary to compute a syntactic analysis of the source string. This is the task of parsing, described in 3.3.1.4, and it is solved by processing the information encoded in the source string together with given, general information about the source language system. From the perspective of computing the translation, it is not necessary to process all information available

[28] Cf. the discussion in Dyvik (1999: 229–230).

about the words in the source string; what is needed is sufficient morphological and syntactic information in order to identify all lexemes, here including function words, contained in the source string and to derive the constituent structure of the source string. Analysing the semantic structure of the source expression is, for instance, not necessary; all that is required for translation is the information that in the type of construction found in the source string, there is syntactic equivalence, which within the domain of type 1 correspondences implies semantic equivalence, between source and target string. The result of the analysis step, i.e. task-specific information about the lexemes and the constituent structure of the source string, is, in the subsequent step, the key to identifying the translation task as an instance of type 1.

Next, the subtask of type identification is solved by processing given, general information about the interrelations between source and target language systems. Identifying a translation task as a type 1 case involves checking whether the following two requirements are met. Firstly, every lexical item in the source string must have a target language correspondent with syntactic and semantic properties matching those of the source item. Secondly, in the target language there must be a structure which is equivalent to that of the source string with respect to the linear order of constituents and the assignment of syntactic functions to constituents. In 3.3.1.4 we have explained that prior to the computing of translations, source-target interrelations may be calculated, and information about them may be associated with individual elements of the lexicons and grammars of the two languages, so that the result of the analysis task will provide direct access to information about any TL elements matching the lexemes and structures identified in the source string. Moreover, we argued that through calculating source-target interrelations it is also possible to identify the linguistic properties which are shared by translationally corresponding elements of the two languages. Consequently, when the lexemes of the source string have been identified, it is possible to decide whether each of them has TL correspondents with shared syntactic and semantic properties. Likewise, when the constituent structure of the source string has been derived, it is possible to decide whether it corresponds with a TL structure matching the source structure according to the requirements of type 1 cases. Thus, the outcome of the type identification task is in practice given by the result of the analysis task, and the information needed to solve type identification is the amount of bilingual information present in the constituent structure derived for the source string. If the outcome of type identification is that all type 1 requirements are met, then the translation task conforms with the characteristics of type 1, and the target text can be generated on the basis of information about the lexemes and the constituent structure (phrase-internal structures included) of the source text.

The final step in the translation task, generation of the target string, involves, in a type 1 correspondence, a search for the target language word forms to replace the words of the source string. The sequence of word forms in the target string is already given by the word order of the source string, which is at this point directly accessible from its constituent structure. Information about lexical correspondence relations between SL and TL has already been accessed, and, in the case of lexemes without inflection, this information is sufficient to identify the correct target word forms, but in cases where more than one inflectional form exist further information is required to identify the appropriate word forms.

In cases where the source and target languages instantiate the same morphological categories, such as number on nouns, the target word form is determined on the basis of morphological information already identified in the source text analysis. E.g. since the Norwegian noun form *skjønnhet* in the source sentence (1a), given in 3.3.2, expresses the morphological feature "singular", the English singular form *beauty* is generated.

A different situation holds if there is a conflict between morphological features expressed by an SL word form and features expressed by its TL correspondent. If the consequence is that the two word forms are not semantically equivalent, then the correspondence violates the demands of type 1 on source-target equivalence with respect to linguistic properties.[29] However, it may be allowed within type 1 that corresponding word forms exhibit morphological differences which do not affect denotational properties, i.e. which do not influence the semantic translational

[29] E.g., number differences affect the denotational properties of the corresponding word forms; this is discussed in 6.3.2.1.

properties of the expressions involved. This type of mismatch may be illustrated by gender marking with reference to the language pair Norwegian-Swedish: both languages have obligatory gender marking, and sometimes translationally corresponding nouns of respectively Norwegian and Swedish exhibit different genders. E.g., for the Norwegian neuter noun *skjørt* ('skirt') the linguistically predictable translation into Swedish is the masculine noun *kjol*. If this kind of lexical correspondence occurs in a type 1 correspondence, there is a conflict between the gender information associated with the source string lexeme and that associated with the target language correspondent. The latter piece of information is available after the SL lexeme has been identified and information about its TL correspondent is retrieved from information about the lexical interrelations between source and target language. In such cases the diverging morphological property of the TL lexeme must, for the purpose of generation, overrule the morphological property of the SL lexeme. Due to TL-specific requirements of gender concord, the TL gender marker must also overrule that of the source string if the structure contains any adjectives or determiners governed by the noun in question. Information about such requirements becomes available through analysis of the constituent structure of the source string since the result of the analysis step provides information about the linguistic properties of translationally corresponding elements in the target language.[30]

It should be added that what distinguishes type 1 from type 2 with respect to the generation task is that whereas type 2 requires the retrieval of the corresponding TL syntactic rules, type 1 only requires the determination that such rules exist (cf. 3.3.3.2). The reason is, simply, that when a translation task is identified as a case of type 1, then it follows from the defining criteria of type 1 correspondences that the generation of the translation can be based directly on the constituent structure of the source string.

Thus, type 1 does not necessitate accessing information about the target grammar, although there are some exceptional cases where it may seem necessary to process syntactic information about the target language. To illustrate, the translation of the

[30] This point will be developed below in the discussion of the translation of present tense verbs from Norwegian into English.

Norwegian present tense verb form *har* in (1a) requires, firstly, that the English verb *have* is identified as the translational correspondent of the Norwegian verb *ha*, and, secondly, that the present tense, singular, third person form *has* is chosen among the various inflectional forms of *have*. The Norwegian source word *har* is marked as a present tense form; this provides temporal information to restrict the search to the set of present tense forms, *have* and *has*, in the English verb paradigm. The marking of number and person is obligatory in English present tense verbs, whereas Norwegian verbs are unmarked with respect to both categories. The source word *har* thus carries no information to settle the choice between *have* and *has*. This problem can be solved by using information about the English grammar rule of subject-verb agreement, together with information about the syntactic structure of the source string. In contrast to Norwegian, English requires agreement between the verb and its subject with respect to the grammatical categories of number and person. According to the syntactic structure identified for the source sentence, the subject (*hun*) carries the grammatical features singular and third person, and hence the singular, third person verb form *has* must be chosen in the translation. In this manner this *appears* to be a case of consulting the target language syntactic rules, contrary to the assumptions of type 1.

However, the type of information required here does not really pertain to syntactic structure (i.e. constituent order and hierarchy), but only to constraints among syntactic elements identifiable by function (i.e. the subject and the verb). Hence these constraints can straightforwardly be assumed to have been retrieved in connection with the calculation of source-target interrelations prior to translation (cf. 3.3.1.4, and above): if it is possible to establish a translational correspondence between the specific Norwegian syntax rule which describes the sentence structure of the given source string and an English sentence rule matching the Norwegian one according to the requirements of type 1, then it is also possible to retrieve the information that subject-verb agreement is included in the corresponding English rule and to associate this information with the Norwegian rule. I.e., this information is included in the set of translational properties, with respect to English, which are associated with the Norwegian sentence rule after the interrelations between the two languages have been calculated, and hence it is not necessary to retrieve information about the target rule one more time in order to solve the generation task. We find it motivated to include, within type 1, cases like the given example when the criteria for type 1 are otherwise met.

Thus, the generation step in translation tasks of type 1 requires different linguistic information sources in order to identify the correct target word forms. In general, these sources include correspondence relations between the lexemes of SL and TL, morphological information derived from the word forms of the source string, information about the syntactic structure (which is derived from the source string and, in type 1 correspondences, shared with the target string), and information about morphological restrictions in the target language. All of these sources need not be required in a given string pair.

3.3.2.3 The weight of the translation task in type 1: processing effort

The structure of the translation task in type 1 correspondences, presented in 3.3.2.2, provides the basis for characterising the weight of the translation task in terms of required processing effort. As explained in 3.2.4 and 3.3.1.2, processing effort pertains to finiteness and amount of required effort. Like the description of the structure of the translation task, the analysis of processing effort can be related to the three subtasks, i.e. source text analysis, type identification, and target text generation.

In type 1 correspondences each of these subtasks is assumed to be computable on the basis of the linguistic information encoded in the source string together with the given, general linguistic information discussed in 2.4.2.2. This settles the question of finiteness as these information sources are available in a finite domain, and can be represented in a finite way. What then remains to be considered is the amount of effort required in order to access and process the necessary information.

Firstly, we have seen that source string analysis requires sufficient lexical, morphological, and syntactic information about the source language to identify all lexemes in the source string, and to derive its constituent structure. This is the task of parsing the source string, and the processing effort involved in syntactic parsing has already been commented on in 3.3.1.2.

Secondly, with respect to the subtask of type identification we have previously discussed how the necessary information about SL-TL interrelations is directly accessible after the analysis of the source string has been done. As stated in 3.3.2.2, identifying type 1 cases involves checking whether two specific requirements are met. Concerning the first requirement, we assume that the computational effort involved in verifying that each source lexeme has a target language correspondent with matching syntactic and semantic properties is comparable to the effort involved in looking up information in a table.[31] With respect to the second requirement, we also assume that it is not more complex than the task of table lookup to check whether the syntactic structure identified in the source string is associated with information about a translationally corresponding structure in the target language which is equivalent to that of the source string with respect to the assignment of syntactic functions to constituents. Thus, we assume that the subtask of type identification can be solved in linear time, since the required number of calculations is proportional to the size of the translation task, i.e. the length of the input (cf. 3.2.1).

Thirdly, with respect to the subtask of generating the target string, we have seen that in type 1 correspondences this involves identifying the correct target word forms to replace each word form in the source string. In general, due to the characteristics of type 1 correspondences the generation task does not involve the computing of any linguistic structures, since the sequence of target words is identical to that of the source string. Although the identification of correct word forms in the translation may require accessing different types of information, all these types are, as has been argued in 3.3.2.2, directly accessible after analysis has been done. For this reason we assume that the computational complexity of each replacement of a source word with its target correspondent is comparable to that of table lookup, and consequently we assume that also the subtask of generation is solvable in linear time.

[31] The complexity of table lookup is commented on in 3.3.1.4.

3.3.2.4 Summary of type 1 correspondences

Type 1 correspondences represent the lowest degree of translational complexity on the scale ranging from type 1 to 4. Relations of equivalence hold between source and target string on the levels of syntax, semantics, and pragmatics, and these equivalence relations must hold with respect to linguistic properties that influence the meanings of the two strings and which are obligatorily expressed in respectively source and target language. Moreover, there exist implicational relations between these equivalence relations: in a type 1 correspondence syntactic equivalence between source and target string implies semantic equivalence, which again implies pragmatic equivalence.

We assume that translation tasks conforming to the characteristics of type 1 correspondences are computable as they fall within the domain of the linguistically predictable translation tasks. Solving them requires the following information sources: firstly, sufficient information about the source language to identify all lexemes in the source string and to derive its constituent structure; secondly, sufficient information about the interrelations between source and target language to find out that each source string lexeme has a syntactically and semantically matching TL correspondent, and that the source string structure likewise has a match in the target language; thirdly, information about the word order of the source string in order to generate a target string where the sequence of words is identical to that of the source string, and sufficient information about morphological restrictions in cases where the lexical interrelations between SL and TL are not enough to identify the correct word forms in the target string.

With respect to processing effort, we assume, firstly, that all types of information required to solve the translation task can be represented in a finite way. Secondly, we assume that analysing the source string is, in type 1 correspondences, potentially the most resource-intensive part of the translation task since it involves syntactic parsing of a natural language expression, whereas we assume that the processing effort required by, respectively, type identification and the generation of target word forms to be proportional to the size of the translation task. Thus, the latter subtasks are assumed to be solvable in linear time, while we assume analysis to be a heavier task, but due to the tendency to low syntactic complexity in type 1 correspondences, we expect it, normally, to be computationally tractable. The conclusion is that the translational complexity of type 1 cases is basically determined by the complexity of the given parsing task.

3.3.3 Type 2 correspondences

Type 2 correspondences are translationally somewhat more complex than type 1: it is not possible to translate word by word, but the degree of complexity is low enough to allow translation "constituent by constituent", as in examples (2) and (3), previously given in 1.3.1:

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

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

As in the case of type 1 correspondences, string pairs of type 2 are not frequent with respect to the pair of languages English and Norwegian, and, as type 1, it is a phenomenon caused by a high degree of structural relatedness between original and translation.[32]

3.3.3.1 Linguistic characteristics of type 2

As mentioned in 1.3.1 and 3.3, the four types of translational correspondences are related to each other in a hierarchy reflecting an increase, from type 1 to 4, in the degree of translational complexity. A consequence of this hierarchical structure is that once we have described the least complex type in the hierarchy, this description can serve as a basis for characterising the second least complex type. Thus, type 2 correspondences are subject to the same restrictions as those applying to type 1 (cf.

[32] The proportion of type 2 correspondences within the recorded data is given in table 5.1 in 5.2.1.

3.3.2.1), except for two deviations from those constraints: in string pairs of type 2 there may be differences between source and target string with respect to the sequence of constituents, and/or with respect to the occurrence of function words. The first kind of deviation is illustrated by example (2) above: (2a) has a fronted adverbial (*dessuten*), followed by the verb *virket*, and then by the subject *hun*,[33] whereas in (2b) the subject *she* is in the initial position, and followed by the adverbial *also*, and then by the verb *looked*. The second kind of deviation is illustrated by example (3): in (3a) there is no word form matching the definite article *the* in (3b), and this is due to a grammatical difference between English and Norwegian: definiteness in nouns is in English marked by the definite article *the*, while in Norwegian it is marked by noun suffixes, which in the case of singular masculine nouns like *leilighet* ('flat') has the form *-en*. Example (3) is a minimal instance of a type 2 correspondence as the string pair exhibits only one linguistic deviation that violates the requirements of type 1 while being allowed within type 2.

Thus, in type 2 correspondences the structures of respectively source and target are not so similar that there is equivalence, through the entire string pair, with respect to the sequence of translationally corresponding surface word forms. Still, in type 2 cases there is near-equivalence on the level of syntax, and the same syntactic requirement as was described with respect to type 1 must be fulfilled: source and target string have to be equivalent with respect to the assignment of syntactic functions to constituents. Correspondences of types 1 and 2 have in common that they are syntactically congruent in the sense defined by Johansson (2007: 202): "Translations which preserve the syntax of the original are termed *syntactically congruent* translations."

In order to clarify the distinction between types 1 and 2, we will add that in type 2 correspondences every source string lexeme *with semantic content* must have a translational correspondent in the target string which is equivalent to the source lexeme with respect to both lexical category and syntactic function. In this connection the relevant distinction is between lexical words and function words, i.e. between seman-

[33] This is due to the verb-second restriction which applies in Norwegian when a non-subject appears sentence-initially.

tically heavy and semantically light lexemes. The use of function words is predictable from information about the grammatical structure of a language, and the requirements of type 2 correspondences are not violated by source-target deviations with respect to the occurrence of function words. Otherwise, the further requirements described for type 1 correspondences must also be fulfilled in type 2. I.e., the syntactic structures of respectively source and target string have to be equivalent with regard to compositionally derived semantic properties, and the two strings need to be pragmatically equivalent (cf. 3.3.2.1). Type 2 correspondences are, like type 1, included among the linguistically predictable translational correspondences.

In the same way as was described for type 1, we may observe implications between equivalence relations on different linguistic levels: we assume that information about how source and target languages are interrelated includes information about what sets of constructions of the two languages which correspond translationally according to the requirements of type 2. If a string pair is identified as a type 2 correspondence, there is syntactic near-equivalence between source and target string, and within the domain of type 2 correspondences this implies also semantic equivalence, which in turn implies pragmatic equivalence, between the two strings. Like in the case of type 1 correspondences, information about these implications can be exploited to solve the translation task without doing a deep linguistic analysis of the source string.

3.3.3.2 The structure of the translation task in type 2: information sources

Since type 2 correspondences are linguistically predictable, a translation task of type 2 is solvable within the pre-structured domain of linguistic information. The structure of the translation task is similar to that of type 1 correspondences, but somewhat more complex since it involves computing certain minor structural differences between source and target string.

The initial subtask of analysing the source string involves the same kind of parsing task as the analysis step in type 1 correspondences does, and it requires the same types of information as discussed in 3.3.2.2. Hence, the subtask of parsing will not be commented on further, since it does not contribute to distinguishing between the degrees of translational complexity in types 1 and 2, respectively (cf. 3.3.1.2).

With respect to the subtask of type identification, we have previously explained in 3.3.1.4 and 3.3.2.2 that its solution is implicit in the result of the analysis task. Like in the case of type 1 correspondences, the information needed to solve type identification is the amount of bilingual information present in the constituent structure derived for the source string. We will illustrate type identification with reference to string pairs (2) and (3), given in 3.3.3.

In the case of (2), the analysis of the source string will reveal that (2a) is a main clause of indicative form with a fronted adverbial and subject-verb inversion. We assume that the Norwegian sentence rule which accounts for the constituent structure of the source string is associated with the information that in the translationally corresponding English sentence structure, the subject precedes the verb, while the sequence of other constituents matches that of the Norwegian structure. We also assume it to be available information that the assignment of syntactic functions to constituents in the English structure is identical to that of the Norwegian sentence. Furthermore, a result of the analysis step is that each lexeme in the source string is associated with information about the translationally corresponding target language lexemes, and in this case this information will reveal that each lexeme in (2a) is linked to English lexemes with matching syntactic and semantic properties. The conclusion is that the task of translating (2a) into the English sentence (2b) is in accord with the requirements of type 2, as specified in 3.3.3.1.

With respect to string pair (3), the analysis of (3a) will reveal that the Norwegian string can be translated word by word into English except for the noun phrase *leiligheten* ('the flat'). The analysis of (3a) will identify *leiligheten* as a definite NP, and the syntactic rule which accounts for this NP will be associated with the information that in the translationally corresponding structure in English the definite article *the* precedes the noun. Since this is a function word, the correspondence conforms with the requirements of type 2.

As pointed out in 3.3.3.1, (3) is a minimal example of a type 2 correspondence, since it exhibits only one kind of source-target divergence which exceeds the restrictions on type 1 while being allowed within those of type 2. This illustrates the point that in our approach the translational complexity of a given translation task, or in a given string pair, is determined by the degree of complexity of the most complex subpart of the task (cf. 4.3.6.1).

Generation of the target string requires a constituent structure in order to compute the linear sequence of surface word forms — this holds for all four types of translational correspondences. It has previously been explained in 3.3.2.2 that with respect to the subtask of generating the target string, types 1 and 2 differ in the sense that while generation in type 1 cases can be based directly on the constituent structure of the source string, generation in cases of type 2 requires also some processing of syntactic information specific to the target language. But to the extent that syntactic structure is shared between the source string and the corresponding rules of the TL grammar it is unnecessary to derive again syntactic structure already identified by the analysis of the source string.

What must be computed for the purpose of generation is that part of the constituent structure, the subtree, which is specific to the target language. To achieve this, it is necessary to retrieve the information given by the relevant syntactic rule(s) of the target language grammar, and it follows from the analysis task which rule (or rules) it is necessary to access information about. These TL grammar rules also provide the necessary information in cases of source-target divergences concerning the occurrence of function words: either the generation of the target text requires introducing a function word not found in the source string, or a certain function word occurring in the source string is *not* matched by a function word in the target string, and these facts will follow from syntactic information about the target language.

Otherwise, the task of identifying the correct target word forms requires the same kinds of information as are needed in type 1 cases (cf. 3.3.2.2). Given the restrictions on type 2 correspondences, the words in the target string will either be TL-specific function words or words which correspond translationally to the lexemes identified in the source string according to the same restrictions as those applying to lexical correspondences in type 1 cases.

With respect to example (2) above, the generation task involves processing the English sentence rule which is translationally linked with the Norwegian sentence rule describing the constituent structure of (2a), and hence generation necessitates a reordering of the verb and the subject in relation to the sequence found in the source string. Otherwise, the constituent structure does not need to be changed. With respect to example (3), generation requires the processing of the English NP rule which is translationally linked with the Norwegian NP rule describing the definite noun phrase *leiligheten*, and, as pointed out above, this will produce the target expression *the flat*.

The subtask of generation is the point where solving the translation task demands a larger amount of information in cases of type 2 than in those of type 1. In tasks of type 2 the need for information in analysis and type identification is on the same level as in tasks of type 1. With respect to generation, the two types have in common that information about the constituent structure of the source string must be available, but in type 2 cases generation also requires information about how source and target must be structurally different and about how the correct target structure is derived.

3.3.3.3 The weight of the translation task in type 2: processing effort

As previously explained, processing effort concerns the extent to which necessary information sources can be represented in a finite way, and the amount of effort required in order to access and process them (cf. 3.2.4 and 3.3.1.2). Translation tasks of type 2 are, like those of type 1, assumed to be computable on the basis of the linguistic information encoded in the source text together with the general linguistic information sources given prior to translation (cf. 2.4.2.2). As argued in 2.3.2, the latter information sources are available in a finite domain, and can be represented in a finite way, so that in this respect translation tasks of type 2 exhibit the same properties as those of type 1 do.

Also with respect to the amount of effort needed in order to access and process the necessary information sources, the requirements of type 2 are mostly the same as those of type 1, but differ on one point, reflecting how the two types vary with respect to the structure of the translation task. In 3.3.3.2 we have argued that the subtasks of analysis and type identification require the same types and amounts of information in cases of type 2 as in those of type 1, and thus we assume that accessing and processing these information sources requires the same amount of effort in both types. Hence, the effort required by analysis in translation tasks of type 2 is determined by the complexity of the parsing task involved in the derivation of the constituent structure of the source string (cf. 3.3.1.4).

With respect to the subtask of type identification, we assume that the necessary information about SL-TL interrelations is, as in the case of type 1, directly accessible after the analysis of the source string, and, as explained in 3.3.2.3, the computational complexity of each checking operation involved in type identification is comparable to that of table lookup. We thus assume that the subtask of type identification is solvable in linear time, and this is common to translation tasks of respectively types 1 and 2. What distinguishes them at this point is that certain translational properties associated with the lexemes and structures of the source string are of different kinds in the two types, but this difference has no consequences for the effort involved in type identification.

Concerning the subtask of generating the target string, it is explained in 3.3.3.2 that this is the point where translation tasks of type 2 require a larger amount of information than tasks of type 1 do, since it is necessary to change one or more subparts of the constituent structure of the source string into TL-specific structure. Insofar as the structure of the target string is shared by the source string, this structural information is directly accessible once the source string has been parsed, and the effort involved in retrieving it is comparable to that of table lookup. Computing the necessary structural changes in the target string involves processing the information available in the relevant grammar rules of the target language, and then substituting target-specific structure(s), or subtree(s), for certain part(s) of the original constituent structure. The information needed in order to generate the correct target structure is easily, if not directly, accessible, and the complexity of the task is modest. Each such substitution is an isolated step in the sense that it is independent of possible other substitutions within the same translation task, and for that reason we assume that in type 2 also the subtask of generation is solvable in linear time, since the consumption of computational resources is proportional to the number of such substitutions.

When it comes to the task of identifying the correct target word forms in type 2 cases, we have seen that information about TL-specific function words is accessible along with the processing of syntactic information for the purpose of generating TL-specific subtrees. Otherwise, identifying the target word forms requires, as explained in 3.3.3.2, the same information sources as in translation tasks of type 1, which means that the necessary information is directly accessible as a result of the analysis of the source string, and that the complexity of the task is comparable to that of table lookup (cf. 3.3.2.3).

3.3.3.4 Summary of type 2 correspondences

Type 2 correspondences represent the second lowest degree of translational complexity on the scale ranging from type 1 to 4. As in the case of type 1 correspondences, relations of equivalence hold between source and target string on the levels of syntax, semantics, and pragmatics, but with respect to syntactic equivalence, certain divergences are allowed: source and target string may differ with respect to the sequence of constituents, and/or with respect to the occurrence of language-specific function words. These divergences cannot violate the requirements that source and target string must be equivalent with respect to the assignment of syntactic functions to constituents, and that all lexical words in the source string must have a target correspondent of the same category and with the same function. Thus, source and target string are equivalent with respect to linguistic properties that influence the meanings of the two strings and which are obligatorily expressed in respectively source and target language. As in the case of type 1, there exist implicational relations between the equivalence relations: in a type 2 correspondence syntactic near-equivalence between source and target string implies semantic equivalence, which again is taken to imply pragmatic equivalence in the given texts.

We assume that translation tasks conforming to the characteristics of type 2 correspondences are computable since they fall within the domain of the linguistically predictable translation tasks, as tasks of type 1 do. Solving translation tasks of type 2 requires the following information sources: firstly, sufficient information about the source language to parse the source string; secondly, sufficient information about the interrelations between source and target language to find out that each lexical word in the source string has a syntactically and semantically matching TL correspondent, and that the source string structure likewise has a target language match conforming with the type 2 requirements described above; thirdly, information about the constituent structure of the source string and information about how the constituent structure of the target string must be different, and, finally, sufficient information about morphological restrictions in cases where the lexical interrelations between SL and TL are not enough to identify the correct word forms in the target string.

With respect to processing effort, translation tasks of type 2 require, in comparison to type 1, the added effort involved in computing TL-specific subtrees in the constituent structure. Otherwise, tasks of type 2 are quite similar to those of type 1: firstly, the various kinds of information required to solve the translation task can be represented in a finite way, and, secondly, source string analysis is potentially the most resource-intensive part of the translation task since it involves syntactic parsing, whereas the computational complexity of type identification and generation of target word forms (other than TL-specific function words) is assumed to be comparable to that of table lookup. Thus, we again consider type identification and generation to be solvable in linear time, while analysis will require a larger number of calculations, determined by the demands of the parsing stage. With respect to the computing of target-specific subtrees in the constituent structure, we have explained it to be of modest complexity, assuming it to be solvable in linear time (cf. 3.3.3.3).

3.3.4 Type 3 correspondences

Type 3 correspondences constitute the second most complex class of translational correspondences. They represent translation tasks where linguistic divergences between source and target violate the restrictions on types 1 and 2, but where source and target text express the same meaning. The linguistic relation between source and target string involves greater structural differences in type 3 correspondences than in string pairs of the two lower types. With respect to the language pair English-Norwegian, type 3 cases are more frequent than instances of both types 1 and 2.[34] In 1.3.1 we gave the following string pair as an example of a type 3 correspondence:

(4a) Hildegun himlet lidende mot taket og svarte med uforskammet høflighet: (BV)
‘Hildegun rolled-eyes suffering towards ceiling.DEF and answered with brazen politeness’

(4b) Hildegun rolled her eyes in suffering towards the ceiling and answered with brazen politeness.

3.3.4.1 Linguistic characteristics of type 3

The defining characteristic of type 3 correspondences is that they violate one of the restrictions on type 2 correspondences in the following way: in a string pair of type 3 it is the case that for at least one lexical word in one of the strings there is no translational correspondent in the other string of the same category and/or with the same syntactic function as that lexical word. Source-target divergences of this kind will cause greater differences in constituent structure between source and target string than the differences allowed within type 2 correspondences, but they must not violate the requirement of semantic equivalence between original and translation. I.e., source and target string have to be equivalent with regard to the sets of expressed predicates and arguments, and the relations between the predicates and their arguments.[35] On the other hand, predicates and arguments in respectively source and target need not be contributed by translationally corresponding constituents between which there must exist relations of syntactic functional equivalence, which is a requirement of type 2 correspondences. The characteristic of semantic equivalence is shared by string pairs of types 1, 2, and 3, and, in general, this means that in correspondences of these three

[34] The proportion of type 3 correspondences within the recorded data is given in table 5.1 in 5.2.1.

[35] According to Alsina (1996: 4), a predicate “expresses a relation (or relations) among participants; these participants are called the *arguments* of the predicate.” Thus, the argument structure of a predicate is the specification of how a set of arguments is involved in the relation expressed by the predicate. The predicate-argument structure specifies the number and internal ordering of the arguments of the predicate.

types, the same informational content is *linguistically* encoded in the source string, as well as in the target string.[36] This is a central principle of our analytical framework.

As previously described in 1.3.1, example (4) above contains two instances of divergences which violate type 2 requirements while being allowed within type 3. Firstly, the intransitive verb phrase *himlet* in the Norwegian source string (4a) corresponds with the expression *rolled her eyes* in (4b), which consists of a transitive verb phrase and a noun phrase functioning as direct object.[37] Thus, between these translationally corresponding expressions there is a considerable difference with respect to constituent structure although they are equivalent in terms of predicate-argument structure: the Norwegian verb *himle* ('roll one's eyes') describes the activity of rolling the eyes of the agent, and this is lexically encoded information, so that the existence of the referent of the English object NP *her eyes* is implied by the Norwegian expression. Secondly, the adverb phrase *lidende* ('suffering') in the Norwegian sentence corresponds translationally with the preposition phrase *in suffering* in the English sentence. These phrases share the function of verb-phrase modification, but belong to different syntactic categories. They are semantically equivalent with respect to the way in which they describe *how* the activity of eye-rolling is performed. We may observe that the remaining parts of the string pair (4), i.e. *og svarte med uforskammet høflighet — and answered with brazen politeness*, is a word-by-word correspondence of the lowest degree of translationally complexity, but, as previously noted in 3.3.3.2, the classification of a given string pair is determined by the most complex subpart(s) (cf. 4.3.6.1).

To sum up, in translational correspondences of type 3 we do not find, as in types 1 and 2, syntactic functional equivalence between source and target string. But in order to fall within type 3, source and target string must be equivalent with respect to compositionally derived semantic properties, and in the given texts they must be pragmatically equivalent in the sense of being used to perform corresponding

[36] In chapter 6 we will discuss some minor exceptions to this. These are cases of predictable differences between source and target string in the amount of grammatically expressed information; cf. 6.3.1.1–2.

[37] Cf. 1.4.2.3, where this example is mentioned in connection with mapping problems for automatic translation. The correspondence between *himlet* and *rolled her eyes* can be described as an instance of conflational divergence between SL and TL.

pragmatic functions, or speech acts. The structural divergences between source and target text in type 3 correspondences show that the degree to which there exist implicational relations between equivalence relations on different linguistic levels is smaller in translational correspondences of type 3 than in those of lower types. The information that there is a type 3 correspondence between a source string and a given target string entails that the strings are structurally different, but semantically equivalent, which in turn implies, within the domain of type 3, that they are also pragmatically equivalent in the given texts. Due to the requirement of semantic equivalence, type 3 correspondences are, like those of types 1 and 2, included among the linguistically predictable translations.

3.3.4.2 The structure of the translation task in type 3: information sources

Since type 3 correspondences are linguistically predictable, a translation task of type 3 is solvable within the pre-structured domain of linguistic information, as it is not necessary to process extra-linguistic information or information from the textual context of the given translation task in order to generate a semantically and pragmatically equivalent target text. Because type 3 correspondences exhibit more complex structural differences between source and target string than correspondences of types 1 and 2 do, solving the translation task in type 3 cases is also more complex than in cases of the lower types. We shall see that in comparison to types 1 and 2, the need for information in type 3 grows with respect to the subtasks of analysis and generation, but not with respect to type identification.

As pointed out in 3.3.1.3, the initial subtask of analysing the source string involves the same kind of parsing task in any type of correspondence. In cases of type 3 the syntactic analysis of the source string thus requires the same types of information as it does in cases of types 1 and 2, i.e. sufficient morphological and syntactic information to identify all lexemes in the source string and to derive its constituent structure.

With respect to the subtask of type identification, the same facts apply in type 3 correspondences as in those of the lower types: the solution to the subtask of type identification is implicit in the result of the analysis task (cf. 3.3.1.4 and 3.3.2.2), and the information needed to solve type identification is the amount of bilingual information associated with the constituent structure derived for the source string.

Identification of a type 3 case may be illustrated by string pair (4), shown in 3.3.4. After parsing, the translational properties associated with the lexemes and constituents identified in the Norwegian sentence (4a) will reveal that the English translation (4b) corresponds with the source string according to the requirements of type 3, but not according to those of types 1 and 2, and this is due to two facts. Firstly, the Norwegian intransitive verb *himle* corresponds translationally with the syntactically complex expression *roll one's eyes*, which means that in the translation there will be (at least) two lexical words, *roll* and *eye*, which do not have correspondents in the source text. Still, there is semantic equivalence between the translationally corresponding expressions since the predicate-argument structure encoded in *roll one's eyes*, including the relation of possession expressed by *one's*, is incorporated in the lexical content of the verb *himle* (cf. 3.3.4.1). Secondly, the Norwegian present participle *lidende*, functioning as an adverbial modifying the verb phrase, corresponds translationally with the preposition phrase *in suffering* which is semantically equivalent to *lidende* in the target text. In this case the corresponding expressions carry the same syntactic function, but as the target expression contains a lexical word, the preposition *in*, which has no match in the source expression, the requirements of types 1 and 2 are nevertheless violated.

When a translation task has been identified as a type 3 case, it is necessary to derive information about the semantic structure of the source text in order to compute the target string. I.e., a semantic representation of the source string must be produced, and this is derived compositionally from the syntactic representation together with semantic information associated with the lexemes identified in the source string. The derivation of a semantic representation requires information about the constituent structure of the source string, about the assignment of syntactic functions to constituents, and about any components of meaning encoded linguistically in the source text (e.g. predicate-argument relations, spatial and temporal relations). It is rather a question of definition whether the semantic analysis of the source string should be regarded as part of the analysis step, or as the first step in the subtask of generation since the output of semantic analysis will serve as the input to generation. What is more important is that the solution of translation tasks of type 3 — in contrast to those of the lower types — requires a semantic analysis of the source string because of structural divergences between source and target string. Since the derivation of a semantic representation of the source string is a kind of linguistic analysis, we prefer to regard this step as part of the subtask of analysis.

We assume that the generation of the target string in cases of type 3 must be based on information about the semantic structure of the source string because type 3 correspondences involve structural source-target divergences of a kind that is qualitatively different from those found in type 2.[38] When generation is based on a semantic representation of the source string, the meaning components identified in that representation provide the information required to retrieve from the target language description the specific lexical units and grammatical structures needed in order to generate the given target string.

It is too simple to view the generation task as parsing in reverse. According to Vander Linden (2000: 765), we may regard the *goal* of natural language generation (NLG) "as the inverse of that of natural language understanding (NLU) in that NLG maps from meaning to text, while NLU maps from text to meaning." However, with respect to *methods*, Vander Linden (2000: 765–766) points out important differences between the two types of processes. Firstly, while there is great variation among generation systems with respect to the nature of the input, systems for NLU (including parsers) receive linguistic input only, which is "governed by relatively common grammatical rules" (2000: 765). Secondly, since NLU aims at interpreting natural language input, "[its] dominant concerns include ambiguity, under-specification, and ill-formed input", matters which are not so relevant in NLG because "[the] non-linguistic representations input to an NLG system tend to be relatively unambiguous, well-specified, and well-formed" (2000: 766). Hence, Vander Linden concludes that "the dominant concern of NLG is *choice*" (2000: 766),

[38] Notably, there is no relation of syntactic functional equivalence between the entire source and target strings.

i.e. choosing the best way of expressing the meaning which is input to a generation system for a given natural language.

Thus, the task of generating a target string from a semantic representation of the source string is, above all, a task of making choices, and the search space for these choices is the entire lexicon and grammar representing the target language. Lexical units and grammatical structures are not selected independently of each other in natural language generation, as there are always close interconnections between meaning and structure in linguistic expressions: "In practice (and perhaps in theory too), it is not possible to separate cleanly selection of lexical items and commitments to particular grammatical organizations" (Bateman and Zock 2003: 289). Anyway, the purpose behind the selection is to extract elements of the TL lexicon and grammar in order to cover all of, but no more than, the components of meaning contained in the semantic representation of the source text. In this manner, the semantic representation, i.e. the input to generation, provides the information needed to carry out the target text generation.

In order to make the choices required to cover the input as precisely as possible, it is a precondition that the input is sufficiently specific. The latter is a challenge to be faced by the design of the framework used for semantic representation, as the framework must facilitate the required level of specificity in the representation of the meaning expressed by the source string. In the case of string pair (4), the semantic representation must provide the information needed to select the lexemes behind the word forms in the target string (4b). Moreover, the semantic representation must provide information specific enough to contribute to the identification of correct target word forms. In contrast to the identification of target word forms in types 1 and 2, information about morphological features identified in the source string word forms is not exploited when word forms are selected for the generation of target word forms in translation tasks of type 3, since the selection must be done by combining the information contained in the semantic representation with TL-specific lexical and grammatical information. As far as other details in the generation task are concerned (e.g. the determination of surface word order), they will be dependent on the design of the generation algorithm, and we do not want to discuss further such issues of implementation.[39]

3.3.4.3 The weight of the translation task in type 3: processing effort

As before, processing effort is considered in terms of the extent to which necessary information sources can be represented in a finite way, and the amount of effort required in order to access and process them. Like translation tasks of types 1 and 2, we assume that those of type 3 are computable on the basis of the linguistic information encoded in the source string together with the general linguistic information sources given prior to translation (cf. 2.4.2.2). As previously argued, these information sources are available in a finite domain, and can be represented in a finite way.

With respect to the amount of effort required in order to access and process the required information sources, there are marked differences between, on the one hand, translation tasks of type 3, and, on the other hand, those of types 1 and 2, and the differences pertain to the subtasks of analysis and generation. As explained in 3.3.4.2, in cases of type 3 the need for information in the subtask of type identification is the same as in cases of the lower types, and the required information, i.e. about SL-TL interrelations, is, as in the case of types 1 and 2, directly accessible through information associated with the syntactic representation produced by parsing the source string. As previously argued, we thus regard this subtask to be solvable in linear time.

Then, with respect to the subtask of analysis, it is noticeably more demanding to access the required information in type 3 cases than in translation tasks of the lower types. Firstly, this is due to the fact that analysis in type 3 involves not only syntactic parsing, but also a semantic analysis of the source string, which makes it necessary to process a larger amount of the source language information available prior to translation, as all semantic information given about the lexemes of the source string must be analysed. Performing a full linguistic analysis of the source string in type 3 cases demands a far greater number of calculations in order to retrieve all necessary

[39] For further information see Vander Linden (2000), Bateman and Zock (2003).

information than the number of calculations required to perform the more shallow syntactic analyses which are sufficient in cases of types 1 and 2.

Also with respect to the subtask of generation, it is far more demanding computationally to access the required information in type 3 cases than in translation tasks of the lower types. In 3.3.4.2 we observed that generating the target string from the semantic representation of the source string involves a number of choices for which the search space is the entire TL language description. Firstly, choices such as selecting the appropriate lexemes, grammatical constructions, and inflectional word forms must be settled by processing information contained in the TL language description together with the information available in the semantic representation. Secondly, a given semantic representation may correspond with not one, but a set of linguistic expressions. A natural language will normally offer more than one way of encoding the same informational content. E.g., we assume that the passive sentence *The boy was given a book by the girl* is semantically equivalent with each of the two active sentences *The girl gave the boy a book* and *The girl gave a book to the boy*. With respect to example (4), we suggest that a semantically equivalent alternative to target string (4b) could be *In a suffering manner Hildegun rolled her eyes towards the ceiling and answered with impudent politeness*. Hence, generating a specific target string from a semantic representation of the source string will frequently involve the problem of choosing the most appropriate target string among a set of alternatives which all correspond with the semantic representation, and this selection task can add to the overall complexity of the translation task, unless a random choice is made.

Altogether, the generation task in type 3 correspondences is clearly a resource-intensive problem. It falls outside the present project to describe in detail the computational complexity of natural language generation from semantic representations, but a few general aspects may be noted. If we assume semantic equivalence between a semantic representation and a corresponding linguistic expression, then the generation task may be regarded as a special case of what Shieber (1993) calls the *problem of logical-form equivalence*. Shieber uses the term *logical form* to refer to any kind of non-linguistic representation of meaning serving as input to a natural language generator (1993: 181). Different logical forms may correspond with one and the same syntactic form (i.e. surface expression) in a language, and hence Shieber argues that a language generator must be able to decide whether two logical forms "mean the same" (1993: 180), i.e. whether one of them may be translated into the other, and this is the problem of logical-form equivalence. It is the view of Shieber (1993) that the problem of logical-form equivalence is computable, but probably intractable due to the lack of an efficient solution algorithm.[40] Since we assume semantic equivalence between a semantic representation and a corresponding linguistic expression, we regard generation from a semantic representation as a special case of translating one logical form into another, semantically equivalent, logical form. For this reason we will suggest that the computational complexity of the generation task in type 3 correspondences is in the same class as that of logical-form equivalence. That is, the generation of the target text probably belongs to the set of intractable problems, and will in general be a more computationally demanding task than parsing is.[41] We will not pursue the topic of natural language generation further; it is also still a field with many unanswered research questions.

3.3.4.4 Summary of type 3 correspondences

Type 3 correspondences represent the second highest degree of translational complexity on the scale ranging from type 1 to 4. Relations of equivalence hold between source and target string on the levels of semantics and pragmatics. Implicational relations between such equivalence relations exist to a lesser degree than in the cases of types 1 and 2: in type 3 correspondences there is not syntactic equivalence between the entire source and target strings, but there is semantic equivalence, which implies pragmatic equivalence in the given texts. Source and target string are structurally different in the sense that for at least one lexical word in one of the strings, there is no translational correspondent in the other string of the same category and/or with the same syntactic function as that lexical word. Correspondences of

[40] The notion of an 'efficient solution algorithm' is described in 3.2.1.
[41] Given a sufficiently high degree of syntactic complexity, the task of parsing a natural language expression may also be intractable.

types 1, 2, and 3 have in common that source and target string are semantically equivalent in the sense that the same informational content is linguistically encoded in both of them.

We assume that translation tasks conforming to the characteristics of type 3 correspondences are computable since they fall within the domain of the linguistically predictable translation tasks, as tasks of types 1 and 2 do. The requirement of compositional semantic equivalence between source and target string in type 3 correspondences will be refined in chapter 6, where we will discuss certain cases of predictable semantic differences between translationally corresponding strings.

Solving translation tasks of type 3 requires the following information sources: firstly, sufficient information about the source language to identify all lexemes in the source string, to derive its constituent structure, and to derive a semantic representation containing all relevant components of meaning expressed by the source string; secondly, sufficient information about the interrelations between source and target language to find out that the target string is structurally different in the sense described above; thirdly, sufficient lexical, morphological, syntactic, and semantic information about the target language in order to generate a target string on the basis of the semantic representation of the source string.

With respect to processing effort, we assume that type 3 has in common with types 1 and 2 that all kinds of information required to solve the translation task can be represented in a finite way. Further, we assume for type 3, as for types 1 and 2, that the subtask of type identification is solvable in linear time.

Concerning the subtask of analysis, we have previously observed in 3.3.1.4 that the complexity of syntactic parsing is in principle the same for all types of translational correspondences. The added processing requirement of type 3, compared with types 1 and 2, is caused by the need for a semantic analysis of the source string.

Finally, with respect to the subtask of generation, we assume that translation tasks of type 3 differ sharply from those of the lower types in the sense that whereas a modest processing effort is required by target string generation in cases of types 1 and 2, generation from semantic representations in type 3 is very resource-intensive, in computational terms probably belonging to the set of intractable problems.

3.3.5 Type 4 correspondences

Type 4 correspondences constitute the most complex class of translational correspondences in our hierarchy of correspondence types. They represent translation tasks where linguistic divergences between source and target violate the restrictions on types 1, 2, and 3: in type 4 correspondences there are not only structural, but also semantic, differences between source and target string. With respect to the language pair English-Norwegian, type 4 cases represent the most frequent class of translational correspondences.[42] In 1.3.1 we gave string pair (5) as an example of a type 4 correspondence, pointing out the semantic difference between the translationally corresponding expressions *for å gå i melkebutikken eller til bakeren* ('to go to the milk shop or to the baker') and *to go and buy milk or bread*.

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

(5b) They could talk here without being called in to go and buy milk or bread.

As explained in 1.3.1, and further discussed in 2.4.2.1, the italicised expressions denote different activities, but it may be inferred from background information about the world that both activities can have the same result, and hence (5b) may be chosen as an appropriate translation of (5a).

3.3.5.1 Linguistic characteristics: type 4 correspondences are different

In contrast to the correspondence types of lower translational complexity it is not possible to describe specific linguistic characteristics of type 4 cases. Rather, the class is negatively defined: in correspondences of type 4 source and target string are not equivalent with respect to constituent structure as in type 1 cases, or they are not equivalent with respect to the assignment of syntactic functions to constituents as in

[42] The proportion of type 4 correspondences within the recorded data is given in table 5.1 in 5.2.1.

type 2 cases, or they are not equivalent with respect to compositional semantic properties. As pointed out in Thunes (1998: 29–30), pragmatic equivalence may hold between source and target string in a type 4 correspondence, but not necessarily. Example (6) may illustrate absence of pragmatic equivalence between original and translation:

(6a) …'har du nå vært på et av disse foredragene igjen.' (EFH)
'Have you now been at one of these lectures.DEF again?'
(6b) 'Have you been to one of those lectures again?'

String pair (6) is almost a minimal case of type 4: it is nearly a word-by-word correspondence, but in the pragmatic particle *nå* in (6a) has no translational match in (6b).[43] The Norwegian adverb *nå* ('now') functions in (6a) as a pragmatic particle creating the impression that the speaker probably disapproves of something the addressee has done. The lack of a corresponding expression in (6b) means, firstly, that a certain semantic component present in the source text is not contained in the target text, and, secondly, that the pragmatic effect created by the particle *nå* is not present in the target text. Because of this semantic difference the pair of sentences (6a) and (6b) is categorised as a type 4 correspondence. This example also illustrates the point that even if two corresponding strings are, by and large, structurally equivalent, the correspondence between them is nevertheless of type 4 if it exhibits some semantic divergence.

Moreover, (6) is an instance of pragmatic non-equivalence where we assume that pragmatic equivalence could have been achieved by choosing a different target text. There may be still other cases where original and translation are pragmatically non-equivalent due to cultural divergences between SL and TL. In such cases cultural differences between the two language communities may have the consequence that some semantic content encoded in the source language cannot be matched by a target

[43] There is also a semantic deviation between the translationally corresponding demonstrative determiners *disse* and *those*. The former expresses proximity, and the latter distance.

language expression with a corresponding communicative effect, and it may even be impossible to find a TL expression with matching semantic content.

Here we shall not pursue the factors governing pragmatic equivalence, since our focus is on the fact that type 4 correspondences distinguish themselves through non-predictable semantic deviation between source and target text, which means that the target string does not belong to the LPT set of the source string. The most notable difference between, on the one hand, correspondence type 4 and, on the other hand, types 1, 2, and 3 is that the semantic representation of the source expression is shared by the target expression in correspondences of types 1–3, but not in those of type 4. Also, within the domain of type 4 correspondences there do not exist, as in string pairs of the lower types, any implicational relations between equivalence relations on different linguistic levels.

3.3.5.2 The structure of the translation task in type 4: information sources

We assume that translation tasks of type 4 differ principally from those of types 1–3 in the sense that tasks of type 4 are not solvable within the pre-structured domain of linguistic information, and the need for information required to translate is larger in type 4 correspondences than in any of the other types. The growth in required information, in comparison to the lower types 1–3, concerns notably the subtasks of analysis and generation.

By definition, translation tasks of type 4 are non-computable since they require information types not included in the pre-structured domain of linguistic information. In 2.3.2 we have argued that there is no principle for delimiting a representation of the information sources lying outside the pre-structured domain, granted that our scope is the translation of general language, and not translation within a restricted semantic area. Thus, there is no principled limit on the amount and types of information that could be needed to solve a task of type 4. We regard type 4 cases as those that demand human translation: the human translator is normally capable of collecting as much information as the task requires, for instance by considering a wider textual context, by looking up more background information of various kinds, or by asking other translators for help, so that a target text can eventually be produced. In this manner translation tasks of type 4 are seen as translatable, although they are not computable (cf. 3.2.5).

With respect to the lower types of translational correspondence, we have discussed the structure of the translation task in relation to the assumption that the task is computable. If we consider type 4 tasks to be non-computable, but solvable by humans, it may seem odd to describe the translation task in terms similar to those applied to the computable tasks. On the other hand, since we do not study the human translation process, the descriptive approach is not altered in relation to the structure of type 4 tasks, and the discussion will focus on the aspects that make type 4 cases fall outside the computable domain.

In order to solve the initial subtask of analysis, a type 4 case requires the same kinds of linguistic information as are required in type 3 (cf. 3.3.4.2) in order to derive a constituent structure as well as a semantic representation of the source string. But in type 4 cases, solving the translation task demands an understanding of the source string which goes beyond a syntactic and semantic analysis, and which requires sources of information included neither in the pre-structured domain of linguistic information nor in the information that is explicitly encoded in the linguistic form of the source string.

It is an important aspect of type 4 that on the basis of the source string alone, it cannot be predicted exactly which additional information sources that are necessary. To mention some possibilities, the required additional sources may include general background information about the world, domain-specific technical information, task-specific linguistic information about reference relations, as well as task-specific extra-linguistic information about the utterance situation of the source text, and about the described situation of the source text. With reference to example (5) above, we have previously pointed out in 2.4.2.1 that general background information about the world is required to fully interpret the Norwegian expression *for å gå i melkebutikken eller til bakeren*. At this point we do not want to illustrate further the types of additional information sources required in type 4 cases, as examples of them will be discussed in chapter 6.

As regards the subtask of type identification, we again want to assume that its solution is implicit in the result of the analysis task. In cases of types 1–3 the analysis yields information about the translational properties, with respect to the target language, of the linguistic items identified in the source string. In translation tasks of type 4, either the translator has chosen a target string deviating semantically from the source string although a literal translation (cf. 2.3.3) could have been produced, or the analysis will reveal that for at least some subpart of the source string there is no linguistically predictable correspondence in the specific target string.

Concerning the subtask of target text generation, we assume that required information sources are the semantic structure of the source string, together with information about the semantic deviation between source and target, as well as information about the grammar and lexicon of the target language. In addition, the generation task requires one or more of the information sources mentioned above in connection with the subtask of analysis. As in type 3 cases, the generation task is first and foremost a question of selecting the appropriate lexemes and structures for the target string. In type 3 this is done by choosing elements of the TL lexicon and grammar in order to cover all of, but no more than, the components of meaning contained in the semantic representation of the source string. In type 4 additional information must contribute to deciding which of those semantic components of the source string that are expressed in the target string, and which are not — as well as which components, if any, that are expressed instead. With reference to example (5) again, we assume that inferencing about the situation described by the Norwegian expression *for å gå i melkebutikken eller til bakeren*, together with access to background information about the world, makes the translator choose the semantically non-equivalent English target expression *to go and buy milk or bread*. As in the analysis subtask, it cannot be predicted, on the basis of the source string alone, which additional information sources are required in order to generate the target string. In general, that question is determined by what we will regard as *parole*-related factors, i.e. factors existing either in a wider textual context, or in the extra-linguistic context, or in both.

3.3.5.3 The weight of the translation task in type 4: processing effort

Since translation tasks of type 4 are non-computable, such tasks are, in computational terms, even harder problems than the intractable ones.[44] Like in the case of the information structure of type 4 tasks, the most relevant topic concerning the necessary processing effort is our assumption that the subtasks of analysis and generation require access to additional information not available in the finite domain of linguistic information sources. Type identification demands no more effort than in the lower correspondence types.

With respect to the subtasks of analysis and generation, we may observe that since some of the information needed to translate in a type 4 correspondence is not included in the finite domain of pre-structured linguistic information, there is in principle no limit on the size of the search job involved in compiling the necessary information. Moreover, within the present approach to translational complexity we have no framework for describing the amount of computational resources needed to access and process such additional information. As stated in 3.3.5.2, we consider type 4 tasks to be cases where human translation is needed, and how hard or easy it is for a human to solve a given translation task will be dependent on that individual's translator competence. That topic falls outside the scope of the present study.

3.3.5.4 Summary of type 4 correspondences

In contrast to translation tasks of the lower types 1–3, type 4 tasks are non-computable as they do not belong to the domain of linguistically predictable translation tasks. Type 4 correspondences represent the highest degree of translational complexity on the scale ranging from type 1 to 4. There is not semantic equivalence between the entire source and target strings; pragmatic equivalence may exist, but not necessarily. Hence, there do not exist, as in string pairs of the lower types, any implicational relations between equivalence relations on different linguistic levels. Type 4 correspondences typically exhibit structural divergences between original and translation, although in certain cases these may be of a minimal kind, as illustrated in

[44] Intractable problems may be computable; cf. 3.2.1.

3.3.5.1. The defining characteristic of type 4 is non-predictable semantic deviation between source and target text, which means that the target string does not belong to the LPT set of the source string. It also means that there will be certain semantic components which are not shared between the semantic representations of each of the two strings.

Solving a translation task of type 4 requires, like tasks of the lower types, access to the information linguistically encoded in the source text, as well as to general, given information about SL and TL, and their interrelations. But producing the semantically non-equivalent target text requires access to additional information sources in order to understand the source string beyond the levels of syntactic and semantic structure. As explained in 3.3.5.2, these additional sources are not included in the finite domain of linguistic information.

Since it is necessary, in order to solve translation tasks of type 4, to access information sources falling outside the finite, pre-structured domain, there is in principle no limit on the processing effort required to search for the needed information.

3.4 Summary

The present chapter is divided into two main parts. The first part (3.2 with subsections) provides a theoretical background for our approach to measuring translational complexity, and the second part (3.3 with subsections) contains a detailed description of the correspondence type hierarchy.

The main purpose of 3.2 with subsections is to view the notion of complexity from different angles, i.e. those of information-theory, linguistics, and translation, respectively, in order to explicate the approach taken to translational complexity in the present approach. In 3.2.1 we start by defining ‘computability’ as a property of tasks: a task that can be solved by a specifiable procedure is a computable task. Further, ‘complexity’ is a mathematical property describing the amount of time and space needed to solve a computable task, and computational complexity theory offers tools for analysing the information processing structure of computable tasks (or problems), as well as for sorting such problems into classes according to complexity measurements. Section 3.2.2 describes the relevance of complexity theory to studies of natural language: firstly, applying complexity analysis to natural language problems has provided new knowledge about the structure of such problems, and, secondly, complexity analysis has proved to be a useful tool in the development of grammar formalisms. However, as pointed out in 3.2.3, applying complexity theory in linguistics requires that natural language problems are studied as problems of computation. Hence, there are several researchers, e.g. Dahl (2004) and Miestamo (2006), who study complexity in natural languages without using the tools of complexity theory. A common denominator of these two contributions is that complexity in natural languages is measured in terms of the length of the description of a given linguistic phenomenon.

Section 3.2.4 discusses the present approach to translational complexity. Like Dahl (2004) and Miestamo (2006), we have chosen a quantifiable, objective commodity as a basis for the analysis of complexity, but whereas their measurement is tied to the length of the description of linguistic phenomena, our measurement concerns the information needed in translation. Moreover, our approach resembles computational complexity analysis in several ways. For one thing, as computability is a precondition for complexity analysis, so translatability is a prerequisite for describing translational complexity. Further, in our analysis translational correspondences are viewed as tasks to be solved, and the complexity of these tasks is described in terms of the structure and weight of a given task. Also, our sorting of translational correspondences into types according to the degree of translational complexity resembles the sorting of computational problems into complexity classes. Section 3.2.5 stresses the point that when analysing translational complexity our focus is on isolating the computable, i.e. the linguistically predictable part of the translational relation, within the translatable.

The presentation of the correspondence type hierarchy in 3.3 with subsections is intended as a way of describing the information situation of the translation task: which information sources are available, and how much information is needed in order to solve a specific translation task? The description is thus an attempt at abstracting away from the human translator in order to investigate to what extent specific bodies of parallel texts could have been translated automatically. Chapter 4 presents the empirical method applied in this investigation.

The four correspondence types differ in several respects. They are described in terms of the linguistic characteristics of the relation between source and target string, and with respect to the structure and weight of the translation task. The structure of a translation task pertains to the amounts and types of information required to solve it, and its weight concerns the effort needed to process those information sources. We keep the subtask of source text disambiguation apart from the translation task, which means that the complexity analysis of given translational correspondences applies to source strings, given their relevant interpretation, in relation to specific target strings.

The correspondence types are organised in a hierarchy reflecting a gradual increase in the degree of translational complexity: type 1 is the least complex, and type 4 is the most complex. Along this hierarchy we may observe, firstly, an increase with respect to linguistic divergence between source and target string; secondly, an increase in the need for information and in the amount of effort required to translate, and, thirdly, a decrease in the extent to which there exist implications between relations of source-target equivalence at different linguistic levels.

Different levels of interpretation of the source string are required in order to solve the translation task. With respect to types 1 and 2, it is not necessary to analyse the source string further than to the levels of constituent and functional structure, as there is a high degree of structural similarity between source and target string. Thus, the target string may be generated on the basis of information about the structure of the source string, and about lexical and structural correspondences between SL and TL. With respect to type 3, it is necessary to derive also a semantic representation of the source string, since source and target string exhibit structural divergences, although they are semantically equivalent. The target string may be generated on the basis of the information contained in the semantic representation of the source string, and the information available in the TL grammar and lexicon.

With respect to processing effort, the workload required by syntactic parsing of the source expression is shared by all correspondence types. In types 1 and 2, since the target expression can be generated mainly on the basis of the constituent structure of the source expression, the major effort of the translation task is involved in the syntactic analysis of the source text. Moreover, in the language pair English-Norwegian, types 1 and 2 tend to occur in correspondences between relatively short and syntactically simple expressions, so that, altogether, a modest processing effort is required by translation tasks of these types. Then, in type 3, the translation task is far heavier than in the two lower types, firstly, since the subtask of analysis requires a full semantic analysis of the source expression, and secondly, because it is a resource-intensive computationally problem to generate the target expression on the basis of a semantic representation of the source string. In the case of type 4, analysing the source string to the level of semantic structure is no longer sufficient, as there is semantic divergence between original and translation, and it is necessary to exploit additional information not explicitly encoded in the source string, nor available through given, general linguistic information, in order to interpret the source string and generate the given target string. Hence, we assume that translation tasks of type 4 are non-computable as they fall outside the linguistically predictable part of the translational relation, and, given our framework, there is in principle no limit on the processing effort that may be required to solve them. On the other hand, tasks of types 1–3 are computable, and their solutions are predictable from the linguistic information contained in the source expression and in the finite domain of information about the two language systems.

According to the view taken in 2.3.3, the notion of literal translation covers correspondences of types 1–3. The most important distinction drawn in the present framework is the division between, on the one hand, the computable correspondences of types 1, 2, and 3, and, on the other hand, the non-computable correspondences of type 4. The computable, or linguistically predictable correspondences have in common that source and target expression are semantically equivalent in the sense that the same informational content is linguistically encoded in both of them. The importance of the analytical distinction between the computable and the non-computable will become clear through the discussion of empirical results in chapter 5.

Non-literal translation, represented by type 4 correspondences, can be seen as the topic of studies of human translation, and in a sense falls outside of the present project. However, in order to clarify the division between the computable and the non-computable, we find it useful to discuss both literal and non-literal translation. For that reason we will in chapter 6 present certain linguistic phenomena which seem impossible, or at least very difficult, to include in literal translation.