

Codified Close Translation as a Standard for MT

Lars Ahrenberg

Department of Computer and Information Science
Linköpings universitet, SE – 58183 Linköping, Sweden
lah@ida.liu.se

Abstract. This paper proposes that quality criteria are set up for machine translation systems that attempt to capture what can be considered the upper limit for the performance of MT systems. It argues that the best performance MT currently should aim for is close translation, and that attempts should be made to codify what that entails for any given language pair. Some guidelines for the selection of close translations are proposed and their application on concrete examples drawn from an English-Swedish parallel corpus is discussed. The definition of different quality levels in relation to a high-quality standard is also touched upon.

1. Introduction

Machine translation research and development have long since abandoned the goals of "Fully automatic high quality" translation. While this is obviously a realistic decision, one may wonder if anything has come to replace it. Practical utility is perhaps the most obvious answer and experience tells us that MT systems of very different quality and coverage may be used to good advantage.

Evaluation is another aspect that has come to the fore as a prerogative for advances in machine translation. Coupled with practical utility evaluation tends to focus on factors that are relevant for the purpose at hand and translation quality may not be the most prominent factor then. However, disregarding the price factor, there can be no doubt that systems are generally more useful the larger their coverage is and the higher their accuracy.

As regards translation quality, evaluation refers to criteria such as fidelity, intelligibility and fluency (White, 2003). These aspects are usually measured by comparisons with reference translations produced by human translators, whether by human judges or automatically (Papineni et al., 2001). While the reference translations are usually good translations, or even "expert translations", their qualities, or the requirements given to the

translators, are seldom discussed in any detail, at least not publicly. Similarly, parallel corpora, such as the Hansards, that are used for training statistical MT systems, are produced by human translators aiming at high quality by human standards. But this quality level may actually be beyond reach for any known system and translation approach developed to date.

The question is, then, what output quality we can expect and demand from a MT system, today and in the future. This is an important question both for researchers, potential customers, and the society at large. Researchers would benefit from having challenging though realistic goals to reach. Customers and users would benefit from having standards to compare with. In particular, it would be useful for customers to have access to quality labels that would give rough but reliable information on the linguistic scope and quality of the output from the system (cf. Hutchins, 2000). Finally, language communities would be better off if the production of mistranslations and gibberish could be kept to a minimum.

In this paper I make a proposal for how the question of goals and standards in terms of translation quality can be approached. Simply put, the proposal is that we should use available resources, such as contrastive grammars and, in particular the parallel corpora that are now being created and annotated in large numbers, in combination with our knowledge of what

makes machine translation hard, to specify the highest quality requirements that we can expect a system to meet. I will argue below that these requirements would fall within the bounds of what is often called close translation. And since detailed specification is needed, I call it a proposal for codifying close translation. The proposal also entails that we can define quality levels by reducing the requirements on the system and use these to rate and communicate the properties of a given system. The IAMT Certification Group has defined three levels in terms of qualities such as dictionary size, sentence types the system can translate, user facilities and intended use (Hutchins, 2000). An assumption of this proposal is that distinctions relating to grammar coverage would also be helpful.

It might be said that all (high-quality) MT systems developed to date have been developed to meet specified requirements. If so, however, these requirements are not known, and it is also not known how well the end product meets them. It is an important part of this proposal that a code for close translation is made public and preferably developed as a community effort.

A code for close translation primarily deals with syntactic constructions and the grammatical words and morphemes that go with them. Of course it is of the utmost importance that a system has a large lexicon and can handle the ambiguities of content words including multi-word predicates. This is a separate quality aspect, however. Another complementary quality aspect is accuracy. Obviously, a close translation need not be free of errors. Table 1 displays a cross-tabulation of three quality criteria that are of primary importance in characterising the properties of a MT system: coverage of the source language, coverage of translation possibilities and accuracy. In this paper the focus is on the second aspect as it applies to grammar, though several comments will be made on its relation to the other qualities, since they cannot be treated independently.

	SL Coverage	Translation Coverage	Accuracy
Grammar		X	
Lexis			

Table 1: Translation quality criteria and the scope of this paper

Issues of quality cannot be discussed very deeply in the abstract. Thus, it is necessary to look into the relations between specific languages. Moreover, it is to be expected that translation quality can be higher the more closely related the languages under consideration are, and this would apply to all linguistic levels from orthographic conventions to pragmatics. In the paper I will draw examples from a current project of mine in making a proposal for a codified close translation manual for the language pair English-Swedish.

2. Modes of translation

The field of translation studies has identified a number of modes of translation that differ in the degree to which there is a correspondence between translation and original. The number of modes may differ between writers but the general idea is to identify a few landmarks as we go from one extreme to another. At one end we have translations that give a counterpart in the target language to every word, or sometimes even every morpheme, of the source sentence, and perform no structural changes whatsoever. This is known as word-by-word translation, or in the most extreme case, as morphematic translation, and is of little use outside the fields of linguistics and language teaching.

At the other end we find translations that represent the content of the source text in a manner appropriate for a particular target audience. At this end, often called adaptation, adherence to structure is not a primary concern and there may even be many changes in content and lexis due to cultural, economic or judicial factors.

2.1. Human modes

In between the extremes we find the dichotomies of formal vs. dynamic equivalence

(Nida & Taber, 1969), or semantic vs. communicative translation (Newmark, 1988), where the one member emphasizes the requirement that a translation should represent the content and style of the original, and the other member emphasizes the requirement that a text should communicate a message to a reader in the most transparent way. Finer distinctions are possible to make, however, and Newmark (*ibid.*) provides a hierarchy of eight categories to cover the whole space. Here I will discuss three categories that Newmark calls literal, faithful, and semantic, respectively, though I use the term ‘close’ rather than ‘faithful’.

A literal translation differs from a word-for-word translation by adhering to target language grammar. However, words are translated out of context and all kinds of figurative or metaphorical uses of language are missed, as the word is regarded as the only semantic unit of interest. Thus, literal translation produces texts that are non-idiomatic and often with a strange or even funny character.

A close translation is one which renders the source text as completely as possible using target language grammar and paying regard to the textual function of words, not only their concrete meaning. A close translation pays attention to structure and does not paraphrase. Correspondences should be found at the lowest level possible; at phrase level if not at word level, at clause level if not at phrase level, and so on.

A semantic (or flexible) translation is like a close translation in that it should give a complete rendering of the contents of the source text. However, it pays more attention to the reader's ability to receive the content, and the fluency and naturalness of the target text. In a semantic translation the structure of the source is less important than the style and aesthetic value and so the language of a semantic translation is more varied than that of a close translation.

2.2. Machine translation modes

MT has been described as a special mode of translation (Sager, 1994). One aspect of this is that automatic translations tend to be structurally close to the source structure, and in

any case, much closer in terms of structure than human translations (Ahrenberg & Merkel, 2000). Another aspect is that the number of alternatives offered for given source language words and constructions is generally smaller than what a translator can produce without much effort. While some approaches such as example-based MT or interlingual MT can cope better with structural differences and functional variants than other approaches, it is generally true that the more variation, the harder it is for the system to pick one which is appropriate in a given context. This is true also for statistical systems that generally perform better if source and target can be made more structurally similar before training starts (cf. e.g. Nießen and Ney, 2004).

2.3. Close translation as a goal

Given this state of affairs it seems reasonable that MT should aim for reaching as far as possible towards the semantic end of the literal-semantic continuum. As a first step, I propose that the goal be set to close translation. This would be a desirable goal for a high-quality gisting system or a core system on which to develop domain-restricted systems.

If this can be agreed, it should have some important effects on both training and evaluation, since the data one should use for training and evaluation should have been translated according to the requirements of close translation. This, in turn, means that those requirements need to be specified. Such a specification is what I'd like to call codified close translation, since it needs to be detailed, not just a list of general descriptions with a few illustrative examples. At the same time, it cannot be fully formalized either, since this would make the task too hard and the community too small.

A close translation can be described as one whose parts can be aligned exhaustively with corresponding parts of the source language. We may call this The Alignment Criterion which in more precise terms can be spelled out as follows: (1) Every clause of the source will have a counter-part in the target; conversely, every clause of the target will have a counter-part in the source; (2) Syntactic phrases and word tokens of the source, with only few and

systematic exceptions, have at least one exponent token in the target; conversely every syntactic phrase and word token of the target must, with only few and systematic exceptions, be an exponent of at least one source phrase or token; (3) Text level alignments belong to types, that can be instantiated in a variety of contexts.

3. Codifying close translation

It is a fact that a translation is underdetermined by the source language text. For this reason some authors have argued for interactive architectures where the system asks the user for help whenever the information in the SL text is insufficient to determine a safe translation (e.g., Johnson & Whitelock, 1989; Somers et al., 1990), and others have argued that relevant world knowledge should be modelled to support decisions (e.g., Nirenburg et al. 1991).

It may well be that the goals for MT, when developed for general (unknown) text, will have to be set at a low level, lower than this paper assumes. However, it is still of interest to know where that level is, and how it can be characterised in the terms of translation studies, given that we have access to more and more translation data to inform system development.

3.1. General selection criteria

The Alignment Criterion in itself allows a wide range of possible translations. For this reason it would be valuable to have guidelines available as we approach the task of selecting translation solutions. An overall goal is that the chosen solutions together should cover the space of possibilities as far as possible. At the same time, their conditions of application should be clearly identifiable so that translation errors are avoided. However, this goal is hard to achieve, and the error rate of a set of options is hard to estimate without empirical testing. The following criteria for selection of useful correspondences seem to be strong candidates, however:

Semantic equivalence. A target language item that consistently has the same or a similar meaning potential as a given source item, is likely to be more useful than one that requires contextual support to convey a similar content.

Structural similarity. An item of the translation that has the same structure as the corresponding item of the source is generally easier to produce for any MT system;

Absolute frequency. The conditions of occurrence for correspondences that occur in high numbers are usually easier to identify and describe, and, can be estimated statistically with greater confidence;

Relative frequency. A choice made more often spontaneously by human translators is likely to be more natural and expected than one which is used relatively rarely;

A problem is, of course, that these criteria often point in different directions. A typical case is the English preposition 'of' when used as a genitive. There are three common ways of rendering that meaning in Swedish, as illustrated below:

E: the roof of the house

S1: taket på huset

"the roof on the house"

S2: husets tak "the house's roof"

S3: hustaket "the house roof"

If a preposition is used in the Swedish translation it will be one with a more specific meaning than 'of'. Thus the choice of preposition in a Swedish phrase of the form 'NP1 P NP2' corresponding to English 'NP1 of NP2' will depend on knowledge about likely actual relations between referents of NP1 and NP2, which makes the task difficult.

The s-suffix underspecifies the relation in quite the same way as the preposition 'of'. On the other hand it induces a structural change and puts more demands on the system for this reason.

The third possibility, compounding, is generally more underspecified than the s-suffix and may lead to misinterpretations, although it is sometimes the preferred choice:

E: the turn of the century

S: sekelskiftet (not 'seklets skifte')

E: the flat of his hand

S: handflatan (not 'handens flata')

E: *platitudes of the media*

S: *mediaplattityder*

(here 'mediernas plattityder' would be ok)

One may regard the first two examples as lexicalized, but since the option is there one would like to have it under control. It is also not so easy always to distinguish 'of' with a genitive reading from other readings. A few examples of this kind are

E: *the study of literature*

S: *litteraturstudiet*

E: *the quality of life*

S: *livskvalitet*

The fact that general criteria do not suffice for the task of selection makes it necessary to study different items in detail and see whether it is actually possible to determine the conditions under which the different translations can be used. Ideally, one would like to arrive at one of the following situations: (1) Under a specified set of conditions, only one option is possible and is thus obligatory; (2) Several options exist, but one of them can be used as a default and the use of the others can be attributed to special conditions, or be found only for lexicalized forms; (3) Several options exist and can be used interchangeably. Often, however, we will find that several non-equivalent options exist, although it is quite difficult to specify the conditions that favour the use of one over the others. This seems to be the case with the *of*-genitive, although my tentative decision is to regard the Swedish *s*-genitive as the default translation. Generally speaking, this is the kind of situation where we have to make practical decisions on the basis of frequency of use, or observed error rates (and call for more research).

3.2. Towards a descriptive framework

While the primary aim of the code is to determine the translation options for an MT system, we may start out from a more comprehensive description that covers the variation we can find in human translations.

From this comprehensive grammar we can then select those options that best meet our requirements (as defined in the previous section). The selection process can proceed as follows:

- What constructions of the source language should be included?
- Which of the different translation options for a given construction constitute an optimal set?

It should be noted that the two languages have different roles. We are not putting them on an equal footing but rather try to describe what happens when one language is the source and the other the target. Thus, entries are taken from the source language. An entry covers a set of SL strings with associated translations. In referring to the SL and TL parts we use labels in the form of more or less elaborated grammatical descriptions.

The pair of an entry and one of its translation options forms a relation that in principle may have an infinite number of instances in parallel texts. Thus, the code provides a partition of possible alignments at the string level, at the same time providing each partition with a reference in grammatical terms. When stating the conditions for this relation, we often need to refer to linguistic material in the surrounding context, which may be small or large. In the descriptions I use the attribute 'parameters' to refer to the relevant contextual material, and the attribute 'scope' to refer to the size of the context.

Salkoff (1999) codifies the relation between French and English constructions using schemata, where a schema is, basically, a string of words and category symbols, where the symbols in turn may represent a cluster of syntactic properties, functional role and, possibly, a semantic category. Similar schemata are also used in the proposed framework, but they are complemented with other types of information. In particular, we need attributes that relate entries to each other. Different entries may apply to the same strings and in these cases we should give information as to which entry (if any) takes precedence, or whether they are in competition. Also, the

status of an option as obligatory or optional is important information.

Another difference between Salkoff's framework and mine is that he seems to look at his rules as recipes for translation. The translation relations codified in this framework may be looked at that way also, but they need not. The idea is not to prescribe the internal workings of an MT system, but to define the possible end results.

Table 2 gives an overview of the most important attributes of the framework.

Attributes	Explanation
Reference	A label for the SL entry in grammatical terms
Level	Word, phrase, clause, ...
Scope	Word, phrase, clause, ...
Forms	A list of instances. The list is marked as exhaustive, finite or open-ended.
Schema	A structure referring to relevant linguistic parameters and constraints.
Superior entries	Entries that take precedence over the current one
Competing entries	Entries that may apply to the same forms and source schema
TL option	A label in grammatical terms for a class of translations
Status	The use of a TL option as obligatory, default, optional or as an exception.
Proportions	A percentage indicating how common the TL option is in a given corpus.
Example	Source forms with associated translations for a given option

Table 2. A list of attributes for coding source language items and their translation options.

3.3. Examples

This section gives some concrete examples of how translation relations can be coded. To save space, not all of the attributes are used, nor are the descriptions always complete. Moreover, for each case we discuss how the selection criteria should be applied.

3.3.1. A function word: The definite article

The English definite article is the most common word token of the English language. Swedish also has definite articles, but unlike French or German, also use a definite noun suffix with approximately the same function. In some contexts these two exponents for definiteness are used in combination, the so called double definite, in other contexts only one of them is used, and, not infrequently, none of them is. In those cases the noun phrase usually contains a determiner other than the definite article. Altogether this amounts to four different translation options that are shown in Table 3.

In addition to the options displayed in Table 3 we find examples where the translator has used a different determiner. It is even possible to use an indefinite Swedish NP to translate an English definite NP.

E: *The rapid and efficient processing of petitions is therefore an excellent means of increasing people's confidence.*

S: *En snabb och effektiv behandling av petitioner är därför ett utmärkt sätt att öka människornas förtroende.*

In general, Swedish often accepts both indefinite and definite articles for generic references. While a definite NP in the translation (*den snabba och effektiva behandlingen*) would suggest a specific reference, a translation without any article would work equally well. However, this would still require the adjectives to be in indefinite form. Unless the conditions for the choice of the indefinite article can be specified clearly, this may be a case where we would prefer an MT system to perform sub-optimally.

Another case of explicitation occurs with references using family names. A noun phrase such as *the Weasleys* can be translated simply as *Weasleys*, but in many cases this would sound insufficient, and the translator would use an elaborated phrase such as *familjen Weasley* (*the Weasley family*) or *bröderna Weasley* (*the Weasley brothers*) instead. Again, to perform such a feat, the system would need an ability to understand references that may go beyond what is currently achievable.

SL reference	The definite article
Forms	(exhaustive) The, the
Schema	[D X* N Y*], where D is the entry, N is the head of D, X and Y.
Related entries	Comparative determiner
Option 1	Double definite
Forms	(finite) (<i>den+ -en</i>), (<i>det+ -et</i>) ...
Schema	[D' Z+ N' W*] where Z translates X or Y. ¹
Example	<i>the big box</i> : <i>den stora lådan</i>
Option 2	definite noun suffix
Forms	(finite) -en, -n, -et, -t, -na
Schema	[N' Z*]
Example	<i>the box</i> : <i>lådan</i>
Example	<i>the letter to Mary</i> : <i>brevet till Mary</i>
Example	<i>the fire alarm</i> : <i>brandlarmet</i>
Option 3	definite/determinate article
Forms	<i>den, det, de</i>
Schema	[D' Z* N' W] where W is a relative clause
Status	optional
Example	<i>the mistake that he made</i> : <i>det misstag han gjorde</i>
Option 4	Null translation ²
Schema	[G Z* N' W*] where G is a genitive noun translating an of-genitive Y.
Status	Obligatory
Example	<i>the capital of Sweden</i> : <i>Sveriges huvudstad</i>
Schema	[Z N' W*] where Z translates X or Y with an adjectival determiner such as <i>samma, nästa, följande, ...</i>
Status	Default
Example	<i>the following day</i> : <i>följande dag</i>

Table 3: Swedish translation options for the English definite article.

¹ An accent on a symbol means that it corresponds in translation to the unaccented symbol.

² This class falls into a number of sub-classes of which we only mention a few here. A list of nine such sub-classes can be found in Svartvik & Sager (1977).

3.3.2. Parts-of-speech: Adjectives

For all of the common parts-of-speech in translating from English to Swedish, the default case would be a translation of the same part-of-speech. So, a noun would be translated by a noun, an adjective by an adjective, and so on. This is not always the case, however, and so the description should account for whatever regularities there are when a translation using a different part-of-speech is chosen.

Some English adjectives are most naturally translated by verbs. An example:

E: *It is not necessary.*

S: *Det behövs inte.*

Gloss: It needs not (It is not needed).

This case is best handled by regarding the relevant construction to be the combination of the copula and the adjective, as in Table 4.

Reference	Adj-predication
Level	Word level
Scope	Verb phrase, Clause
Schema	[B Y* A Z*], where B is a copula, and A is an adjective
Option 1	A-to-A mapping
Schema	[B' W* A' U*]
Status	Default
Option 2	A-to-V mapping
Schema	[V W*] where V translates B and A
Status	Lexical exceptions

Table 4: Swedish translation options for predicative adjectives.

Translating an attributive adjective by a relative clause is not uncommon. This seems to happen mostly for morphologically complex adjectives, where Swedish does not have a corresponding lexical form, or when modelling a Swedish adjective on the English construction would yield awkward results:

E: *unanswerable*

S: *som inte gick att besvara*

Gloss: that were not possible to answer

E: *lightning-shaped*

S: *som liknade en blix*

Gloss: which resembled a flash

E: *heart-stopping*

S: *som fick hjärtat att stanna*

Gloss: that made the heart stop

This type of translation seems to be a challenge for current systems, but not completely beyond state-of-the-art (cf. 3.3.4).

3.3.3. Word order differences

While the previous examples have treated specific linguistic material in translation, there are order differences between English and Swedish that are so general as to apply to any constituent that happens to be in a position that is simply not available in a Swedish sentence. While English readily allows both a subject and an adverbial phrase before the finite verb of a main clause, Swedish does not. The normal solution for the translator is to keep either the subject or the adverbial in the first position and express the other in an appropriate position to the right of the finite verb. Which one is placed first varies with a number of factors such as which one comes first in the source, the kind of adverbial concerned, and, probably, general discourse considerations. It is beyond the scope of this paper to give a detailed analysis of the problem, but in Table 5 I give an example to illustrate the format of rules that primarily record word order differences.

3.3.4. Fluency vs. Intelligibility

In making selections we have referred to structural similarity as one of the selection criteria. Different kinds of structural shifts differ in their degree of difficulty, however. The most difficult are the ones where the system is required to insert linguistic material, that have no simple counterpart in the source. One such example was mentioned in the discussion about adjectives in 3.3.2. One may then say that the alignment criterion is not fulfilled, but as we allow (systematic) exceptions to it, we are faced with the problem of drawing the line on a case by case basis. Now, English and Swedish are quite similar as regards grammatical distinctions, but there are some cases where

Reference	V3 main clause order
Level	Phrase level.
Scope	Declarative main clause.
Schema	[A N V X*] where N is the subject of V, V is finite, A is an adverbial and X is in the same clause as V.
Option 1	Inverted V2-clause
Schema	[A' W N' Y*], where W is finite and translates V or X.
Status	Optional
Example	<i>In spite of their repeated requests nothing has happened : Trots deras upprepade efterfrågningar har ingenting hänt.</i>

Table 5: Incomplete account of the translation of English main clauses with the finite verb in third position.

human translators add linguistic material quite regularly. One such case is the use of non-finite verbs as clausal heads. This is much more common in English than in Swedish, so translators tend to use a finite clause instead, which entails that a a tense, a subject, and, often, a subjunction or a modal, is part of the translation. Purposive infinitival clauses constitute one example:

E: *To view total or detail data, ...*

S: *Om du vill se sammanlagda data eller detaljdata*

Gloss: If you want to view

A common translation of 'to' is the Swedish conjunction 'för att'. It could be used here with the result of an intelligible sentence, but not one which would be regarded as good style, unless the clause is moved to an internal sentence position. This would then entail a substantial structural difference between source and translation.

E: *Reading the newspaper last night, I ...*

S: *När jag läste tidningen i går kväll, ...*

Gloss: When I read ...

A literal translation, using the Swedish present participle 'läsande' would yield an ungrammatical sentence, though probably intelligible for a reader with some knowledge of English. Another possibility would be to use a nominalisation such as 'vid läsning av' (at reading of), but this is still awkward and requires the retrieval of an appropriate preposition.

If the non-finite clause can be analysed as being coordinated, the subject may be left unexpressed, but there is still a need to retrieve the tense and the coordination:

E: *He escaped from the Warsaw ghetto, leaving behind his parents and his sister.*

S: *Han flydde från Warsawas ghetto och lämnade kvar sina föräldrar och sin syster.*

Gloss: and left behind

Again, using the Swedish present participle 'lämnande kvar' in this case would be stylistically very marked, if not ungrammatical.

Similarly, when the clause functions as a relative, a relative pronoun and a tense is required in the translation:

E: *And the murderer, speaking to one of the geniuses of the nineteenth century, ...*

S: *Och mördaren, som talade till en av artonhundratalets stora genier ...*

Gloss: who spoke

My inclination is that high-quality English-Swedish MT should aim for the most natural translations in these cases, i.e., for producing finite clauses for a range of English non-finite source clauses. It is not an easy task, however, and many applications might do without them. Thus, it is also an instructive case for seeing the value of being able to grade systems using quality levels.

3.4. Quality levels

A detailed description of the options available for a close translation and for the contexts of their occurrence can provide the basis for the definition of quality levels. A system that covers everything that the code covers, and can apply it with high accuracy and good fluency to

unseen texts, i.e., at the level of a human translator, will undoubtedly be of a standard not yet reached by any system. Still, its flexibility would be much smaller than that of a human translator, since its ability to produce paraphrases and explicitations would still be limited. Similarly, to order clauses and phrases differently from the author of the source text to meet target language requirements on natural discourse would probably be beyond reach.

Lower quality levels can be defined by excluding alternatives that, although they satisfy the criteria of a close translation, do so to a lesser degree, e.g., by being less frequent or structurally more demanding.

Similar quality levels can be defined for the coverage and treatment of content words. For this purpose large existing bilingual dictionaries in combination with monolingual and parallel corpora can provide the basis for defining levels. Pairs of source and target words can be rated on grounds of frequency and generality in much the same way, and possibly, with more ease than pairs of grammatical items.

Generally speaking, the more alternatives a system can handle accurately, the better it is. In defining quality levels we should then attend to factors such as the following:

(i) Lexical coverage; As noted, the IAMT Certification Group has proposed the size of the dictionary, as a relevant attribute. It needs to be explained, though, whether this applies to the source language only, or to the number of different translations the system covers in principle;

(ii) Lexical accuracy; the precision (or word error rate) of the system with respect to the content words that it actually attempts to translate.

(iii) Grammatical coverage; the number of source language constructions it covers and the number of accurate translation options for grammatical words and constructions that the system can in principle provide. The highest to be expected is then full coverage of the code.

(iv) Grammatical precision; the precision (or error rate) in translating grammatical constructions;

It would seem that the procedures for evaluations along these lines are already in place. The major difference would be that

reference translations would be required to follow the code and the parts of it that are included at a given quality level.

4. Summary and outlook

A proposal has been made for codifying a type of translation which should be within reach for state-of-the-art machine translation in the coming five- or ten-year period. I have positioned this type of translation as what is often called close, or faithful, translation, in descriptive translation studies. I have argued that this would give researchers more focused aims and provide a basis for the definition of system quality levels that are more detailed and informative than those currently in use. Thus, it would be possible for users to learn the basic qualities of a system by reading a description of it.

No doubt codifying any type of translation is a substantial undertaking for any given language pair. Moreover, it needs to be done in conjunction with system development. Some general characteristics of what to include or exclude in a code, have been proposed, and other more detailed guidelines have been hinted at. Others are invited to join the discussion.

5. References

Ahrenberg, Lars and Merkel, Magnus (2000). Correspondence measures for machine translation. In Proceedings of the LREC 2000 Workshop on Evaluation of Machine Translation, Athens, Greece 29th May, 2000, pp. 41-46.

Hutchins, John (2000). The IAMT Initiative in Defining Translation System Categories. Fifth EAMT workshop, Ljubljana Slovenia 11-12 May 2000. Accessed from <http://nl.ijs.si/eamt00/>.

Johnson, Roderick L. and Whitelock, Peter (1987). Machine Translation as an Expert Task. In S. Nirenberg (ed.) *Machine Translation: Theoretical and Methodological Issues*. Cambridge University Press, pp. 136-144.

Newmark, Peter (1988). *A Textbook of Translation*. Pearson Education.

Nida, Eugene S. and R. Taber (1969). *The Theory and Practice of Translation*. Leiden: E. J. Brill.

Nießen, Sonja and Ney, Hermann (2004). Statistical Machine Translation with Scarce Resources Using

Morpho-Syntactic Information. *Computational Linguistics*, Volume 30, Number 2, pp. 181-204, June 2004.

Nirenburg, Sergei, Carbonell, Jaime, Tomita, Masaru and K. Goodman (1991). *Machine Translation: A Knowledge-Based Approach*. Morgan Kaufmann.

Papineni, Kishore Salim Roukos, Todd Ward and Wei-Jing Zhu (2001). Bleu: a Method for Automatic Evaluation of Machine Translation. IBM Research Technical Report.

Sager, Juan C. (1994). *Language Engineering and Translation: Consequences of automation*. John Benjamins 1994.

Salkoff, Morris (1999). *A French-English Grammar: A Contrastive Grammar on Translational Principles*. Amsterdam, John Benjamins.

Somers, Harold L. (ed.) (2003). *Computers and Translation: A translator's guide*. Amsterdam, John Benjamins.

Somers, Harold L., Tsujii, Jun-ichi, and Danny Jones (1990). Machine Translation without a Source Text. In Proceedings of the 13th International Conference on Computational Linguistics, Helsinki, 1990.

Jan Svartvik and Olof Sager 1977. *Engelsk universitetsgrammatik*. Almqvist & Wiksell Förlag AB.

John S. White 2003. How to evaluate machine translation. In H. Somers (ed.) 2003: 211-243.

