

Kielikone Machine Translation Technology and Its Perspective on the Economics of Machine Translation

Dr Harri Arnola

Abstract

This paper describes the machine translation technology of Kielikone Ltd. and gives an outline of TranSmart, a Finnish-English system which is a commercial application of that base technology. We argue that MT is fundamentally empirical research. Product development is a slow and strenuous process, and MT systems will remain incomplete not only vis-à-vis human translations but also with respect to the system's own potential translation quality. An evaluation method is described which measures the progress in MT development. This method can also be used for system and technology evaluation. The paper concludes with the claim that the real contribution of MT will be seen in the long term in applications that do not compete with human translations but in which MT is the only option.

Dr Harri Arnola

Harri Arnola teaches Artificial Intelligence in Helsinki University of Technology. He was the leader of the Kielikone project that was funded by the Sitra Foundation, a public Finnish fund whose objective was to support risky R&D that carries obvious national interest. The aim of the project was to design and implement computational algorithms that would be widely applicable language technology applications in Finland.

Kielikone Ltd.

The Kielikone project resulted in the setting-up of Kielikone Ltd., a Finnish language technology company that currently has 13 employees. Harri Arnola is the managing director of the company. Kielikone Ltd. is the market leader in Finland as a supplier of proofreading software and electronic dictionaries. The main R&D activity of Kielikone is Machine Translation technology and MT systems. The **TranSmart** Finnish/English system is the first MT system that is based on Kielikone's own MT technology. It is used by two major Finnish companies. The system has been used also to provide MT services to various companies and individuals. Kielikone's MT work has been supported by the Technology Research Centre of the Ministry of Trade and Industry in Finland.

Kielikone Group

Kielikone Ltd. is the parent company of a group of small companies. The other companies in the group are **Transwise Oy**, a Machine Translation services company, and **Wordmark Oy**, a marketing company, located in Jyväskylä, Finland. In addition, Kielikone Ltd. is a shareholder in **Käännöskone Oy**, a company that specializes in hand-held electronic dictionaries, located in Vihti, Finland.

Harri Arnola

Kielikone Ltd., P.O. Box 126, 00211 Helsinki

Tel: +358 0 6820 211, Fax: +358 0 6820 167

E-mail harri@kielikone.fi

(Authors: Arnola¹, Harri; Hyvönen, Kaarina; Juntunen, Jukka-Pekka; Linnanvirta, Tim; and Suoranta, Petteri of Kielikone Ltd.)

1. KIELIKONE MT

This section describes the base MT technology developed by Kielikone. The technology is language independent and can be used for building MT systems for various language pairs. Kielikone has built a commercial MT system, TranSmart, applying the base technology. The TranSmart system translates from Finnish into English.

1.1 Base technology

In terms of machine translation techniques, Kielikone MT technology (KMTech) uses the transfer approach; as to linguistic theories, KMTech makes a commitment to the dependency theory. Transfer, then, consists of a sequence of meaning-preserving transformations of source dependency trees into target dependency trees.

An **MT Engine** is the basic process for transforming dependency trees for a specific goal. An MT Engine instance is a virtual machine which consists of two parts: the generic code (the engine proper) and a task-specific rule base (Figure 1).

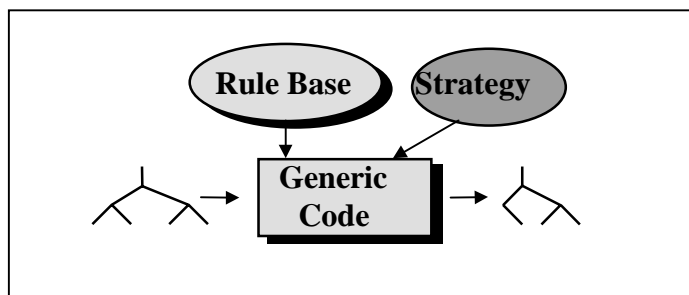


Figure 1: MT Engine

An MT system which uses KMTech has an extremely simple but flexible linear architecture. The transfer is a chain of the MT Engine applications (Figure 2). Each engine performs a certain transfer subtask. In KMTech also lexical transfer is performed by MT Engine applications. The number of MT Engine applications used depends on how the transfer problem is divided into subproblems. The architecture can be changed at any time by defining new subtasks and inserting new MT Engine applications in appropriate places in the chain. For example, an MT system may comprise several distinct lexical transfer phases and their priority is assigned simply by changing their order. A full MT system consists of an input analyzer and a dependency parser, which are not part of KMTech, a KMTech MT Engine transfer chain, a linearization module, and a postprocessing module. The two synthesis modules can also apply the MT Engine (Figure 3).

¹ Formerly Jäppinen

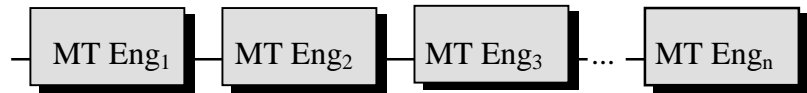


Figure 2: Linear transfer architecture

Fully automatic machine translation has distinct quality limits. One way to overcome such limits is to access and adapt high quality human translations (or corrected MT translations) stored in a translation memory. Obviously, the extent to which translation quality is improved in such a hybrid system depends first and foremost on the amount of recorded data. Since in practice an exactly matching sentence is only rarely found in the memory, the added value of a translation memory depends also on how intelligently the translations of nearly matching sentences can be used. Translation memories usually perform fuzzy matches on **strings** of words. KMTech offers a linguistically intelligent translation memory which stores the dependency **trees** of source sentences. Matching trees rather than strings offers several benefits. To name one, dependency trees neutralize variations in constituent order in a natural way thus making the adaptation of near-matches easier (Juntunen, 1996).

1.2 A case study: the TranSmart Finnish-English system

Kielikone's TranSmart Finnish-English system is a fully implemented MT system which relies on KMTech (Jäppinen et al., 1991, Jäppinen et al., 1993). Its basic architecture is shown in Figure 3. The shaded squares represent MT Engine applications. The analysis has two parts: morphological analysis of word forms, and dependency analysis of sentence structures. Deterministic parsing in linear time is the main theoretical attraction of the parser (Arnola, 1996). As the figure shows, transfer activates the MT Engine seven times (lexical transfer three times). As a product, TranSmart has two versions: a workstation solution (Unix), and a client-server solution (Unix).

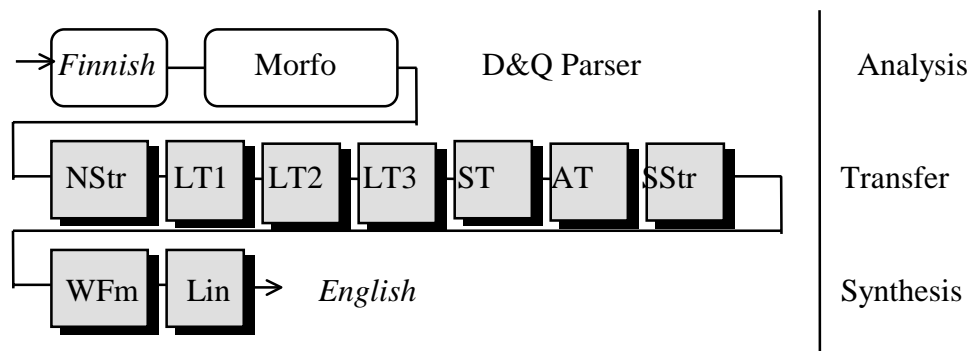


Figure 3: Architecture of the TranSmart Finnish-English system

1.3 TranSmart in practical use

Kielikone MT has been under development for several years. The project was initially supported by the Sitra Foundation, and later on, in the product development phase, by the Technology Development Center of the Ministry of Trade and Industry. From the outset there has been strong user participation in the work. The companies listed below have participated in the development work both financially and concretely. Although systems have already been sold to customers outside this consortium, Kielikone has so far mainly concentrated on catering to the needs of the members of the consortium. Major marketing efforts have been deferred.

The consortium

Nokia Telecommunications, net sales over FIM 10 billion, a subsidiary of the Nokia Group, is one of the pilot customers who have participated in the development work from the beginning. The original intent of the company was to use TranSmart as a workstation to support customer documentation. In this role the translation memory feature plays an important role as documents often have common sections. During the current year (1996) a new application has surfaced spontaneously at Nokia. The company has plenty of employees whose native tongue is not Finnish stationed in various countries. Occasionally they face documents which are written in Finnish. Thanks to the TranSmart server version installed in the company net they are now able to obtain quick rough translations which often satisfy their needs.

Rautaruukki Oy, a Finnish steel company, net sales over FIM 9 billion, is another pilot customer. The TranSmart system is installed as a server in the company net, and employees can access it through the internal electronic mail system.

Trantex Ltd., a Finnish translation services company, net sales over FIM 28 million, is a third pilot customer. The company specializes in localizing software products but produces translation services in other fields as well. Text types and domains tend to vary greatly in a translation services company and since MT requires domain specific lexical tuning, the benefits of MT are not so clear in this area.

MT services

Jointly with Trantex Ltd., Kielikone has established a translation services company, Transwise Oy, to offer machine translation services fast and at a reasonable price. This service function has two goals. The first and paramount goal is to offer customers a new form of attractively priced translation services. The second, subsidiary goal is to collect texts from different sources for purposes of system tuning. Since the texts are used also for tuning, translation speed has not yet been as high as machine translation technology would ordinarily allow. Transwise Ltd. intends to install the TranSmart system in the WWW in the near future and start offering instantaneous translation services.

Next we want to share our experience in the strenuous work of building MT systems. MT systems have idiosyncratic characteristics which often make the work seem frustrating and evasive. In fact, MT systems are never complete. Nevertheless, they do have promising applications.

2. MT Systems Are Incomplete

It is no surprise to anyone that MT systems are incomplete. This section attempts to argue briefly why incompleteness is an essential characteristic of MT systems. This discussion serves as a stepping stone to the discussion on evaluation of incomplete systems.

2.1 *There is no theory of MT*

Incompleteness is a result of the lack of rigorous theories of MT. We might start by asking if there exists a verifiable theory of translation of sentences between two languages in the sense of hard sciences. By such a theory we mean a detailed and explicit method or a set of formulae which relates source language sentences to their perfect translations in the target language, **without resort to human intuition**. We believe that such a theory does not and will not exist, at least in a near future. (Of course there are other forms of theories of translation which teach and study the trade of translation, but these theories are not rigorous in the sense we are looking for.)

One of the reasons why there are no scientifically precise theories of translation is that translation is about meaning-preserving transformations between utterances in two languages, and there is no rigorous theory of meaning available. It is unlikely that anything except fragmentary theories of meaning of natural languages (say, Montague semantics or situation semantics) will surface in the foreseeable future. Furthermore, natural languages abound idiomatic expressions and conventions whose correlations with other languages constitute an empirical and not a theoretical inquiry. In any given natural language there are dozens of idiomatic expressions which resist any precise semantic analysis which would unequivocally link the expressions to another language. Such expressions can be only approximately projected into another language and the projections are language specific.

If our argument is true, it follows that there neither is nor will be a computational theory of translation. A computational theory presumes a conceptual theory; if the latter is missing, so is the former.

2.2 *Yet, MT can and should be theoretical*

However, from the argument it does not follow that MT cannot be **theoretical**. It can (and should) be theoretical, in at least two different ways. First, MT should use linguistic theories of word and sentence structure and text cohesion. There are several morphological theories which relate word forms with their morphological structure and syntactic theories which relate sentences with their syntactic structure. Moreover, there are computational algorithms which produce such theoretical structures more or less reliably. As translation is about meaning-preserving transformation of sentences, it is advantageous, we believe, to choose from among alternative syntactic theories one that indicates functional rather than constituent structures of sentences. Our choice is the dependency theory.

Second, even if there is no comprehensive computational theory of translation worthy of the name, MT should be theoretical in the sense that translations should capture as many linguistic generalizations between two languages as possible. Even if such a set of generalizations falls short of a comprehensive theory, it takes steps towards that direction. And, applying Occam's razor, we might then say that one MT system for a given language pair is more "theoretical" than another if the former applies more

salient generalizations about the language pair than the latter (and presumably needs a smaller number of contrastive rules). From this viewpoint, even word for word translation systems are “theoretical” to an extent (since probably they implement at least some rudimentary generalizations between the language pair), but they are less “theoretical” than more advanced systems.

2.3 Hence, MT systems are incomplete

If there is no comprehensive computational theory of translation, MT technologies, as embodiments of certain theoretical principles for the purpose of producing translations, are incomplete. Thus, if we are right, there cannot be an MT technology which would produce a closed solution for the problem of translating sentences between a given language pair in the sense that one would be able to conceive of a set of formulae that, once written down, would produce perfect translations between the pair. The problem is and remains open. Theoretical research is able to produce a partial solution but there remains a considerable empirical work load. Consequently, the quality of an MT system is not a step function but improves only slowly as new empirical associations are added to the theoretical base (Figure 4). The more “theoretical” a system is, the faster its quality improves and the higher is its asymptotic, potential translation quality. MT systems are not only incomplete vis-à-vis human translation quality; they are also incomplete with respect to their own, system dependent potential translation quality.

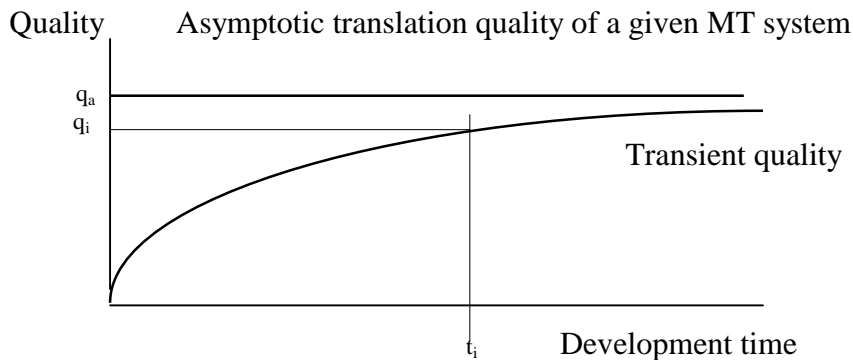


Figure 4: Gradual improvement of the quality of an MT system

MT system development is slow and labor intensive work and its goals are evasive. MT is therefore not only a theoretical endeavor, but it has also clear engineering goals. MT technologies should support both the implementation of linguistic theories and the unavoidable empirical work. KMTech, we believe, is such a technology.

3. Evaluation of MT Systems

Usually artifacts are so designed that they serve the intended purposes fully. MT systems, being inherently incomplete, are therefore odd commodities. The issue of the economics of imperfect commodities, per se, is outside the scope of this paper and we shall content ourselves with only a few remarks on the question.

Incomplete systems have no absolute value; they have only relative (or functional) value. Usually a product, say, a boat, has absolute value in that you can use it for crossing any waterway. An MT system is like a slightly leaking boat in that it cannot be used for all translation needs but it can, nevertheless, be used for some specific purposes. It has therefore only relative value, and one needs evaluation to secure that the quality of a given MT system is high enough for a specific purpose.

A specific purpose is often a certain functional role, say, the production of rough translations for a human translator. An incomplete system qualifies for a functional role only if its use creates added value in the whole process. For example, if an MT system is used for producing rough translations for a human user, the positive value of the system (rough translations) must be greater than the negative value which its use generates (postediting cost). The best way to ensure that an MT system qualifies for a certain functional role is, again, to perform evaluation. The rest of this section discusses evaluation methods of MT systems.

3.1 The purposes of MT evaluation

If incompleteness and far from perfect performance is inherent in a system, evaluation becomes very important. It is not acceptable to market an MT system either by claiming that it is perfect, which is false, or by just claiming that it is imperfect. The latter claim, albeit honest, leaves the customer puzzled. He or she rightly asks, but **how** good is it? To answer such a natural question one needs reliable evaluation.

There are several different purposes one might want to evaluate MT systems for. At least the following purposes have been discussed.

System comparison. A potential user of an MT system for a given language pair wants to compare different systems in order to choose the right one (e.g. Flanagan, 1994).

System evaluation. A potential user of an MT system for a given language pair wants to evaluate a particular system in order to see if it is good enough (in terms of quality and economy) for the intended purpose.

Technology comparison. A funding agency or an MT system developer wants to compare different MT technologies in order to see which one is worth funding or choosing, if any (e.g. White et al., 1994).

Technology evaluation. An MT system developer wants to evaluate a specific MT technology in order to see if it provides a good enough base technology for MT system development.

Progress evaluation. An MT system developer wants to evaluate progress in MT system development work in order to see when the system has reached its limits and the cost of additional work no longer pays off (e.g. Ishara et al., 1994).

3.2 Black box evaluation and glass box evaluation

Of these, system evaluation is by far the most frequently performed one although results are rarely reported. Whenever a potential user thinks seriously of using an MT system in production environment some kind of system evaluation is performed. The

most frequently discussed evaluation issues seem to deal with comparing systems or technologies. Two methods have often been mentioned.

In **black box evaluation** test systems are given identical inputs and their outputs are compared and ranked. If the systems are mature such a strictly behavioristic evaluation method is fair and equitable. For a practical working system it should not matter what internal processing takes place; all that counts is what the system delivers.

If the systems are not mature and it is permitted to correct errors located during evaluation, it makes a difference how the corrections are made. If an error results from a missing piece of linguistic knowledge or a missing or incomplete lexical entry and if the fault can be fixed following the standard procedure used for implementing linguistic knowledge in the system, the error does not indicate a theoretical or structural weakness in the system. If, on the other hand, the fault can be corrected only in an ad hoc manner by associating hand-coded translation with the input or a part of it, the error reveals a fundamental weakness in the system.

In **glass box evaluation** test systems are given identical inputs and their outputs are compared and ranked as in black box evaluation. Furthermore, corrections are allowed and the theoretical principles and linguistic generalizations employed by the systems are observed and ranked. Glass box evaluation is fair if the systems are not mature or if they always require customer specific tuning before production.

Black box and glass box evaluations are plausible choices for system comparison. When technologies are compared, glass box evaluation is preferable to black box evaluation. Next we discuss an evaluation method used by Kielikone. This method, which is in the spirit of glass box, has applications in progress evaluation and system evaluation, and it may offer material for technology evaluation as well.

3.3 Window evaluation

Kielikone developed the method of window evaluation originally for the purpose of monitoring progress in system development work. As argued earlier, the quality of an MT system approaches its final quality extremely slowly. At any give moment t_i there is a “distance” $|t_a - t_i|$ between the current quality of the system (t_i) and its potential quality (t_a) (Figure 4). The distance should decrease monotonously while the work progresses. Window evaluation attempts to give a measure for such a distance and to monitor the monotony of the progress.

For the sake of this discussion, let us assume the concept of a **translation space** generated by various linguistic attributes. A point in the space is said to be **covered** if the system can translate the source structure represented by that point at the asymptotic level. Figure 5 shows a hypothetical translation space drawn in two dimensions. (In practice, such a space would, of course, have many dimensions.) This crude visualization attempts to separate areas covered by structural transfer rules representing linguistic generalizations (shaded irregular areas) and lexical transfer rules representing word-specific translations (shaded circles). Due to the empirical nature of the work, the space will always have uncovered areas (white areas in Figure 5). To evaluate the current quality against a certain text type one takes a representative sample of the text. The piece of text covers a certain fixed area in the translation space. (Represented by a rectangular in Figure 5. In reality, the subspace covered by any piece of text would not be continuous.) The piece of text offers a kind

of a “window” into the translation space, and the finite subspace “seen” through the window can be fully covered. White areas (missing or incorrect structural and lexical translations) are revealed in this subspace and the system can be tuned to its asymptotic level.

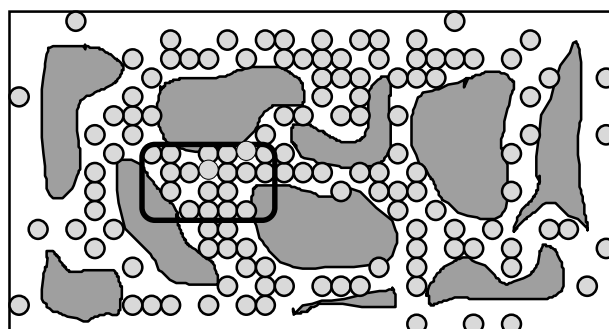


Figure 5: A translation space covered by generic and specific rules

Let us call the initial translation (no corrections) of a piece of text its **rough rough** translation and the final translation (all possible corrections made) its **polished rough** translation. The rough rough translation indicates the current quality, the polished rough translation shows the potential quality, and the corrections made represent the “distance” between the two. Notice that in this context “polished” does not indicate post-editing.

Table 1 gives an example. The text is a light news item printed in a Finnish newspaper. The original text was 56 sentences long. To save space, the table shows only the first 30 sentences. The text has not been pre-edited. Sentence #7 (marked by a star) is parsed incorrectly in two ways: the highly elliptic structure is parsed incorrectly, and a genitive attribute gets an incorrect structure (*Nesteen aurinkoenergiaan...* / *solar energy of Neste* should read *Nesteen... tutkimusosasto / research department of Neste*). The translations of this sentence are therefore quite wrong. Whenever TranSmart encounters a source word for which it has no translation it copies the original source word. These are marked in bold face in the rough rough column. (Headings are also in bold face.) If the missing word is a compound and there exist translations for the parts, TranSmart performs a part-for-part translation as a default. Such translations appear in italics in the rough rough column. The polished rough column indicates systematic corrections between rough rough and polished translations: lexical corrections are in italics and structural corrections are in bold face. Polished rough translations have **not** been post-edited in any way.

#	Source	Rough rough target	Polished rough target
1	AURINKOENERGIAN VOIMALLA LIKKU-VASTA VENEESTÄ PUUTTUU MOOTTORIN	THE SÄKSÄTYS OF THE MOTOR IS MISSING FROM THE BOAT WHICH MOVES WITH THE POWER OF	THE <i>CLATTER</i> OF THE MOTOR IS MISSING FROM THE BOAT WHICH MOVES WITH THE POWER OF

	SÄKSÄTYKS	SOLAR ENERGY	SOLAR ENERGY
2	Kolmen vuoden kehittälyllä yhdistettiin suomalainen puuvene ja aurinkoenergia	With the three year development the Finnish wooden boat and solar energy were connected	With the three year development the Finnish wooden boat and solar energy were combined
3	Tavallisen soutuveneen perässä on pikkuinen moottori.	There is a little motor behind an ordinary rowing-boat.	There is a little motor behind an ordinary rowing-boat.
4	Vene liukuu vedessä kuin unelma, tasaisesti, äänettä-mättä ja saasteettomasti.	The boat slides in water like a dream, evenly, silently and saasteettomasti .	The boat slides in water like a dream, evenly, silently and <i>without pollution</i> .
5	Sähkömoottori saa voimansa auringosta.	The electric motor gets its power from the sun.	The electric motor gets its power from the sun.
6	Suomalaisen soutuvene on työstänyt aurinkoveneeksi Juha Nyman Särkisalosta.	The Finnish rowing-boat has been worked to be a <i>sun boat</i> by Juha Nyman from Särkisalo.	The Finnish rowing-boat has been modified into a <i>solar boat</i> by Juha Nyman from Särkisalo.
7*	Hanketta on ollut kehittä-mässä myös Albican-verkos-toyryitys sekä tukemassa Nes-teen aurinkoenergiaan keskit-tyntyt tutkimusosasto.	The project has also been developed by the Albican <i>network company</i> and has been supporting the <i>research department</i> which has concentrated on the solar energy of Neste.	The project has also been developed by the Albican network company and has been supporting the research department which has concentrated on the solar energy of Neste.
8	"Tämä on kolmen vuoden kehittelyn tulos.	"This is the result of the three year development.	"This is the result of the three year development.
9	Hanke on edennyt pikkuhiljaa ja keväätalvella idea toteutetiin käytännössä.	The project has proceeded little by little and in late winter the idea was carried out in practice.	The project has proceeded little by little and in late winter the idea was carried out in practice.
10	Pääajatus oli yhdistää vanha suomalainen puuvene ja aurinkoenergia", selittää Nyman.	The main idea was to connect an old Finnish wooden boat and solar energy", Nyman explains.	The main idea was to <i>combine</i> an old Finnish wooden boat and solar energy", Nyman explains.
11	Paneeli piiloon penkin alle	Panel to hiding place under bench	Panel to hiding place under bench
12	Teknisesti aurinkovene vaikuttaa sangen yksinkertaiselta.	The <i>sun boat</i> seems very simple technically.	The <i>solar boat</i> seems very simple technically.
13	Aurinkopaneeli lataa akun, josta sähköinen perämoottori saa voimansa.	The solar panel loads the accumulator from which the electric outboard motor gets its power.	The solar panel loads the accumulator from which the electric outboard motor gets its power.
14	Paneelin voi jättää näkyviin tai työntää piiloon istuimen alle. Vene saavuttaa noin kymmenen kilometrin tuntinopeuden ja yksi lataus riittää reilun tunnin ajomatkaan täydellä vauhdilla.	The panel can be left in sight or can be pushed to the hiding place under the seat.	The panel can be left in sight or can be <i>hidden</i> under the seat.
15	Nymanin aurinkovenevers-taaseen on kantautunut uutisia, joiden mukaan Ruotsiin on syntymässä hänelle kilpailija. Muista pohjoismaisista kilpailijoista ei ole tietoa.	The boat will reach about the ten kilometre speed per hour and one charging is enough for the good hour drive on a full speed.	The boat will reach <i>the speed of about ten kilometres per hour</i> and one charging is enough for the good hour drive <i>at full speed</i> .
16	Nymanin aurinkovenevers-taaseen on kantautunut uutisia, joiden mukaan Ruotsiin on syntymässä hänelle kilpailija. Muista pohjoismaisista kilpailijoista ei ole tietoa.	News according to which the competitor being born to Sweden to him have been carried in Nyman's <i>sun boat workshop</i> .	News according to which the competitor <i>is being born in Sweden</i> to him has <i>reached</i> Nyman's <i>solar boat workshop</i> .
17	"Ruotsin TV4 esitteli taannoin ruotsalaisyrityksen valmistaman aurinkoveneen.	There is no information about other Nordic competitors.	There is no information about other Nordic competitors.
18		"Sweden's TV4 demonstrated	"Sweden's TV4 demonstrated

19	Sen sähköratkaisut vaikuttivat tosin kovin alkeellisilta. Myös Saksassa päin on tietävästi muutama tämän-	recently the <i>sun boat</i> made by the Swedish company. However, its electric solutions seemed very elementary. There probably are a few such projects also somewhere in Germany”, Nyman says.	recently the <i>solar boat</i> made by the Swedish company. However, its electric solutions seemed very elementary. Also somewhere in Germany there probably are a few such projects ”, Nyman says.
20	suuntainen hanke”, sanoo Nyman. Tavoitteena Keski-Eurooppa		
21	Aurinkovene on ollut esillä muutamilla venemessuilla.	As objective Central Europe	As objective Central Europe
22	Suomalaiset messuvieraat ovat jonkin verran vieroksu-neet	A few <i>boat exhibitions</i> have had a <i>sun boat</i> up for discussion.	A <i>solar boat</i> has been <i>shown at</i> a few <i>boat exhibitions</i>
23	Nymanin venettä. "Perussuomalainen	The Finnish <i>exhibition guests</i> have shunned Nyman's boat a little.	The Finnish exhibition guests have shunned Nyman's boat a little.
24	soutuveneen ostaja on aika konservatiivinen. Yksi kommentti on kuulunut, että veneemme on digitaalilla pilattu puuvene."	"The <i>basic Finnish</i> rowing-boat buyer is quite conservative.	"The basic Finnish rowing-boat buyer is quite conservative.
25	Sen sijaan keskieurooppalaiset ovat olleet kovasti innostuneita. "Etenkin saksalaiset ovat olleet ihastuneita aurinkoveneen ympäristöystävällisyyteen."	One comment has been that our boat is a wooden boat that has been spoiled with digitaali ." Instead the Central Europeans have been very excited.	One comment has been that our boat is a wooden boat that has been spoiled with <i>digitals</i> ." Instead the Central Europeans have been very excited.
26	Nykyisellään Nymanin vene ei kuitenkaan kelpaa Keski-Eurooppaan.	"Especially the Germans have been attracted to the kindness to the environment of the <i>sun boat</i> ."	"Especially the Germans have been attracted to the kindness to the environment of the <i>solar boat</i> ."
27	Siellä päin veneitä säilytetään pääasiassa telillä, koska vedessä olevia venepaikkoja on niukasti. "Jos nykyisen veneen nostaa aina vesireissun jälkeen telille, puu kuivuu ja seuraa-vassa vesillelaskussa vene täyttyy vedellä." ...(cont.)	As it is now Nyman's boat does not suit to Central Europe, however. There the boats are retained mainly on the spindle because there are scantily <i>boat places</i> in water. "If it lifts the present boat always after the vesireissu to the spindle, the tree will dry and in the following launch the boat will become full of water."(cont.)	As it is now Nyman's <i>boat is not suitable</i> , however, <i>for</i> Central Europe. There the boats are <i>kept</i> mainly on the spindle because there are scantily <i>places for boats</i> in water. "If one lifts the present boat always after the <i>boat trip</i> to the spindle, the tree will dry and in the following launch the boat will become full of water." ...(cont.)
28			
29			
30			

Table 1: Example text in window evaluation

The polished rough translation in Table 1 does not yet represent the final word of TranSmart, since certain phenomena (such as the proper assignment of articles or the proper ordering of adverbials) have not been handled conclusively yet. There are several such errors in the polished rough translations. We already mentioned that translation errors of sentence #7 are mainly due to the parsing errors. Notice how elliptic headings may get awkward translations (#11, #21). In Finnish the word *puu* means tree, wood, or timber, depending on context. In sentence #30 a wrong translation is chosen (*the tree will dry...* should read *wood will dry...*). This error cannot be corrected by the general method and the error is therefore left as its is.

Corrected word specific translations are shown in Tables 2-4. Table 4 does not show semantic information. Nouns can be semantically classified, and this typology can be used in selectional restrictions for example in choosing correct verb configurations.

#	Source	New target
1	säksätys (informal)	clatter
2,10	yhdistää	combine
6,...	aurinkovene	solar boat
15	vauhti(Advl, Ad)	at speed
25	digitaali (rare as a noun)	digital
29	venepaikka	place for boat
30	vesireissu (informal)	boat trip

Table 2: Added or corrected domain specific word or preposition translations

#	Source	New target
4	saasteettomasti	without pollution
27	ympäristöystävällisyys	kindness to the environment

Table 3: Added or corrected generic word or preposition translations

#	Source	New target
6	työstää smth1(Obj) smth2(Advl,Transl)	modify smth1(Obj) into smth2(PComp)
14	työntää smth(Obj) piilo (Advl, Ill)	hide smth(Obj)
15	QuantNoun(QuantAttr) tuntinopeus	the speed of QuantNoun(PComp) per hour
16	olla syntyä (Advl,IIIinf,In) smth(Advl,III)	be (Progr) born in smth(PComp)
16	kantautua smth(Advl,III)	reach smth(Obj)
22	olla esillä (Advl) smth(Advl,All)	show (Pass) in smth(PComp)
28	kelpaa smth(Advl,III)	be suitable for smth(PComp)
29	säilyttää smth1(Obj) smth2(Advl,Ad)	keep smth1(Obj) on smth2(PComp)
32	smth(Advl,All) olla hinta (Subj,Part)	smth(Subj) be expensive (Compl)

Table 4: New or corrected generic configurations

The text revealed also two structural errors. In sentence #20 the surface ordering of adverbials was corrected. In Finnish the passive voice indicates an unknown actor. The passive is usually indicated morphologically with a passive morph in the main verb. An unknown actor can be indicated syntactically using a 3rd person singular finite verb without a subject. Both structures are translated into English using either an English passive structure or using one of the formal subjects *it*, *one*, or *they*. The rough rough translation of sentence #30 opts for the incorrect formal subject *it*.

The distances between the rough rough and the polished rough translations are shown in Table 5. A distance is a real number between 0 and 1. It is calculated simply by dividing the number of sentences requiring corrections by the total number of sentences.

Correction type	Domain specific	Generic	Combined
New words or prepos	4	2	6
Corrected words or prepos	3	0	3
New configurations	0	3	3
Corrected configurations	0	6	6
New structural rules	0	1	1
Corrected structural rules	0	1	1
Lexical distance	0.23	0.33	0.5
Structural distance	0	0.07	0.07

Table 5: Quality distance of the example text

The figures in Table 5 are not informative for the casual reader. For a system engineer they provide relatively straight-forward data on progress. When quality distances are measured regularly, using similar, general text type, the numerical generic distances should demonstrate a decreasing tendency. Since structural distances are further divided into subclasses at Kielikone (see Figure 3), the numbers tell how mature the different transfer parts are and where attention should be focused.

Window evaluation can also serve a potential customer's needs. Both rough rough and polished rough translations are made for a representative piece of the customer's text. When the customer sees both translations and the domain specific distance figures, he or she gets at least an approximate idea about the maturity of the system and its potential performance for his/her text type. The domain specific lexical distance gives an estimate about the amount of lexical work required by the text type before the system can be used in production.

Window evaluation may give data also for technology evaluation. Assume that an MT technology has been used for building an MT system for the language pair a-b and the question is raised how good a quality that technology would provide for another language pair c-d. Polished rough translations show the final quality for the pair a-b for a given text type. If a and b are structurally more distant from each other than c and d, window evaluation tells indirectly how good a quality is achievable for the pair c-d for a similar text type. If, on the other hand, c and d are more distant from each other than a and b, window evaluation remains silent.

Window evaluation does not tell anything about the absolute translation quality of the system. This evaluation method provides only relative figures about the maturity of a given system. The method yields also samples of the final quality, and these samples can be subjected to any of the evaluation methods proposed in the literature of the field in order to get an estimate of the absolute final quality. Informally, these samples tell an informed reader something definite about the quality of the system.

4. Niche Economics of MT

Our discussion has had a slightly negative bend so far since it has mainly dealt with the limitations of MT. This final section straightens things up and ends the discussion with an upbeat note. In search for a better term, we call sudden new economic opportunities opened by the use of a totally new technology **niche economics** (Church and Hovy use the term for a different purpose). A new technology may pave the way for commodities of a totally new kind which consumers have not even

dreamt of. If ordinary product development is driven by demand pull, the products in niche economics, if they take customers by surprise, manifest technology push.

We believe that MT may yield technologies which create their own niche economics. Translation services, as they are currently offered, are greatly limited by the physical limitations of the human body and its central nervous system. Human translations are slow and often inconsistent. A human translator gets easily fatigued and his or her memory is imperfect. Human translation requires the presence of the whole bulky human body and functions only under restricted environmental conditions. Moreover, privacy is breached when a text is translated by another human. Of course this last problem can be legally solved by writing binding agreements, but nevertheless the fact that somebody to whom the text was not intended reads it may in some cases be so great an obstacle that translations are avoided. If, on the other hand, translations are performed by machines, people probably would not feel their privacy threatened.

MT is free of all these drawbacks. It is fast and consistent, a machine never gets fatigued and it has perfect memory. LSI circuits fit in a small space and function in a great variety of environmental conditions. And machines do not breach privacy. But, of course, quality has been quite low at least so far. We believe that the MT technologies already available may be good enough to produce MT products for niche economics. In particular, we believe that KMTech is such a technology.

There are certain products and services which for different reasons are beyond the capabilities of human translators but which MT might deliver in the near future. Interpreting telephone, portable interpreting aid, translation of e-mail messages, translated newspaper, translating fax machine, and translating copying machine are examples of such products or services. As is well known some of these products or services are already under R&D in many countries.

There are probably many more to come that we cannot yet even think of. It seems that discussions of MT economics have often been too preoccupied with short term economics and comparison with human translation. The real strength of MT will be seen in the longer run, in its niche economics, where MT does not compete with humans but turns out to be the only choice.

References

- Arnola, H., On parsing linear dependency structures deterministically in linear time. Manuscript, 1996.
- Church, K., and Hovy, E., Good applications for crummy machine translations.
- Dowty, D., Wall, R., and Peters, S., *Introduction to Montague Semantics*. D. Reidel Publishing Company, 1981.
- Falkedal, K. (Ed.), Proc. of the Evaluators' Forum. ISSCO, 1991.
- Flanagan, M., Error classification for MT system evaluation. Proc. of the First Conference of the Association for Machine Translation in the Americas, 1994.
- Ishara, H., Uchino, H., Ogino, S., Okunishi, T., Kinoshita, S., Shibata, S., Sugio, T., Takayama, Y., Doi, S., Nagano, T., Narita, M., and Nomura, H., Technical evaluation of MT systems from the developer's point of view: exploiting test sets for quality evaluation. Proc. of the First Conference of the Association for Machine Translation in the Americas, 1994.

Juntunen, J-P., Intelligent translation memory. M.Sc. Thesis, Helsinki University of Technology, 1996 (in Finnish).

Jäppinen, H., Kulikov, L., and Ylä-Rotiala, A., KIELIKONE Machine Translation Workstation. Proc. MT Summit III, Washington, DC, 1991.

Jäppinen, H., Hartonen, K., Kulikov, L., Nykänen, A., and Ylä-Rotiala, A., KIELIKONE Machine Translation Workstation. In Nirenburg, S. (Ed.), Progress in Machine Translation. IOS Press, 1993.

White, J., O'Connell, T., and O'Mara, F., The ARPA MT evaluation methodologies: evolution, lessons, and future approaches. Proc. of the First Conference of the Association for Machine Translation in the Americas, 1994.

