The IAMT Certification initiative and defining translation system categories

John Hutchins

1. Background

Since the middle of the 1990s there has been a rapid increase in the number and variety of translation systems available, in the form of stand-alone software for 'automatic' translation, computer-aided translation systems for large corporations, translator workbenches, translation memory systems, on-line systems provided on the Internet (some of them free), and there will no doubt be more in the future. For the general public, computer software for translation is a quite new product; they are unaware of the advantages, limitations and methods of using such systems. They are furthermore familiar with rapid improvements of computer technology and software, and will therefore be expecting similar rapid improvements in the quality of translation software. One of the tasks of the MT community must therefore be to convey some idea of the immense complexities involved in dealing with natural language (and in particular the kind of colloquial language used over the Internet), and to explain why automatic translation has progressed so slowly despite over forty years of intensive research.

For some time the IAMT has been concerned that some guidance should be provided for potential purchasers of systems, to explain, for example, the differences between MT and translation support systems, the differences between systems for 'home use' and systems for large companies, etc. The guidance should also cover the suitability of different types of systems for particular uses and tasks, information about components of systems and how they may be used (including their benefits and limitations), and information for the cost-effective and/or time-saving use of different types of translation support and MT systems. In addition, it is recognized that ideally there should be some authoritative benchmarking or some reliable consumer evaluations, but these activities are currently beyond the resources and competencies of IAMT and its regional associations. It has therefore been decided that as a first step IAMT should attempt to provide a categorization of translation systems together with explanations of what different types of systems can and cannot do.

The IAMT initiative began in 1997 with the Council setting up a small subcommittee under Eduard Hovy (president of AMTA). Initially the idea was that IAMT might establish a »seal of approval« which would be awarded to companies conforming to agreed standards. However, it was quickly realised that this term would imply some evaluation of systems, and this was not intended. Consequently, it was decided that given the limited resources of IAMT as a voluntary organization, it should aim for a more modest approach: a »certification« or categorization of products in a form agreed by vendors, together with some general guidance for potential purchasers.

During 1998 there was an exchange of ideas among those who had expressed interest, and in October 1998, the AMTA conference included a panel where general aims and ideas were aired (Hovy 1998). As a result, the ad hoc group drew up a tentative set of definitions of categories for circulation during 1999 to manufacturers and others for their comments.

In the meantime, work on the »Compendium of translation software« has suggested a slightly different set of definitions which appear to be more easily applied in practice. The »Compendium« has been a parallel, but independent, effort (Hutchins 2000). It is intended as a general guide to commercially available systems, containing only brief information about individual systems and their components. However, in the interests of ease of use, there has been some standardization of terminology, including use of a standard categorization of system types.

In part, the categorization is derived from the IAMT initiative, but a number of changes and additions were made in the light of practical application to existing systems and in accordance with vendors' own descriptions of their systems.

2. Foundations

2.1. Basic terms

It is perhaps to some extent unfortunate that we seem to be saddled with the term »machine translation«. It is seen by many as an old-fashioned term, redolent of a pre-electronic and pre-computer age. More particularly, however, this term implies only purely automatic systems not involving human participation at any stage. It excludes, almost by definition, all kinds of computer aids for translation. What is wanted is an easily understood term that covers computer-based systems that provide translations which can be used as such (without revision) or which can be used as the basis for higher quality human translation, and which covers also many kinds of translation support tools.

The term chosen for the »Compendium« was »translation software«. This is, perhaps, reasonable for covering commercial products (both fully automatic systenms and support tools), but it is not adequate as a term for research and development activity.

An additional area of confusion is the use of the term »computer-aided« or »computer-assisted« translation (CAT). Sometimes it refers to the use of automatic translation systems with facilities for human involvement or intervention, before or after text processing (i.e. pre-editing, controlled input, post-editing). In other words, CAT is used by many vendors in preference to MT. On other occasions CAT refers to computer support tools for translation such as dictionaries, translation memories, etc. There is further confusion from the availability on the market of software combining both types of system, e.g. MT and translation memories.

2.2. Basic distinctions.

The first distinction therefore for the general public has to be between:

a): wholly automatic systems, i.e. systems that (attempt to) translate texts and sentences as wholes

and (b): computer-based translation aids, i.e. systems that provide linguistic aids for translation.

In the latter case, it is easier (and probably clearer) to list the aids, since many of them will already be familiar (even if not in electronic form):

1. Dictionaries: both bilingual and multilingual, with and without grammatical information, with and without guidance on usage (appropriateness)
2. Language aids providing grammatical information (morphology, noun/verb paradigms)
3. Spelling checkers
4. Style checkers
5. Terminology aids, such as glossaries of 'authorized' terminology for a particular scientific, technical or commercial field
6. Specialised glossaries, e.g. for a translator's special subject areas, for particular clients, agencies and customers

Other computer-based aids will not be known by the general public, and would therefore require detailed explanations (see below, section 4). These include tools for pre-editing and controlled language, tools for the creation of corpora of 'approved' translations (translation memories and alignment tools), and management support tools for, e.g. budgeting and cost controls, workflow and personnel management, etc.

3. Automatic translation (MT) systems

3.1. Minimal basic features

Firstly the general public needs to know what distinguishes a 'true' MT system from a 'dictionary translation'. The need for this stems from the existence on the market and on the Internet of systems that are described as translation systems but which are in effect no more than bilingual dictionaries.

As a minimal definition, we can say that a translation program should be more than simply substitute words of the source text by words of the target language. It should provide:

(a) minimally correct morphology. For example, the endings of adjectives should agree with the nouns they modify, the grammatical cases of nouns should agree with the selected verb forms, etc.

(b) some minimal syntactic processing. For example, the order of adjectives and nouns should be inverted when translating between English and French, the position of the verb should be correct when translating between German and English.

(c): some semantic processing. For example, there should be some selection among alternative 'equivalents' according to context or subject field.

In addition, a translation system should permit not only input as phrases or sentences (not just word by word) but it should also produce output in the form of (grammatically) 'complete' sentences.

For the purposes of categorization it is not desirable to give definitions in terms of internal processes (e.g. interlingua, transfer, example-based, statistical analysis, feedback/learning, 'artificial intelligence'). These are relevant when discussing research systems but not for the description of commercial systems and not for the general public.

The definition by the IAMT Certification group is:

> A software system is a »machine translation system« if it takes input in the form of full sentences at a time and generates corresponding full sentences (not necessarily of good quality).

The definition in the »Compendium« is:

> MT system: software for automatic translation, where input units are full sentences of one natural language and the output units are corresponding full sentences of another language.

Both are essentially variants of the same concept, concentrating on input and output as sentences, as opposed to 'word-for-word' dictionary renditions.

The IAMT Certification group has underlined this point by providing a definition for »translation support tools« as a category:

> A software system is a »translation support tool« if it takes input one word or phrase at a time and produces proposed translated equivalents, which users must then assemble into sentences and texts.

In the »Compendium« there is no general definition for translation support tools, instead all the types are listed with their own definitions (see below).

3.2. Levels of utility

It was recognized from the beginning that systems are being developed for users with a wide variety of needs and a wide variety of experiences of translation. It was felt essential to indicate these basic differences in some way. (Although applied to MT the definition of 'levels' might also be relevant for support tools.)

The levels identifiable are:

a) Basic level (or »entry level«, »home use«) covering systems that are not considered suitable for professional translators and not expected to produce top-quality translations. They

tend to be cheap (some are free on the Internet) and are used primarily for information purposes (getting the essence of messages) or for simple correspondence (including electronic mail)

b) »Professional« level (or »standard« level) covering systems intended for professional users (e.g. translators), and while not expected to produce translations of 'publishable' quality without revision, can be used cost-effectively in appropriate circumstances. In general, such systems should provide editing facilities, larger dictionaries than »entry« level systems, and facilities for the creation of user dictionaries.

c) »Company« level (or »advanced« level) for systems intended for large-scale processing (repetitive documents, multiple languages, technical documentation) and which may be integrated into other documentation processes (authoring, controlled languages, publication, etc.). In most cases these systems run on client-server (intranet) configurations.

3.3. Definitions of system types

Currently, the definitions of these levels by the IAMT Certification Group are as follows:

> Machine Translation System (basic level)
>
> A machine translation system is a "Basic level system" if its largest dictionary contains fewer than 50,000 entries OR if has very limited facilities for users to extend the dictionary OR if its translation capability is restricted to the translation of single-clause (basic) sentences. Systems at this level are primarily meant for home use.
>
> Machine Translation System (standard level)
>
> A machine translation system is a "Standard level system" if it contains a dictionary of more than 50,000 root entries AND it is not restricted to the translation of single-clause (basic) sentences AND it provides facilities for the user to extend the dictionary. Systems at this level are primarily meant for home use or standalone office use.
>
> Machine Translation System (advanced level)
>
> A machine translation system is an "Advanced level system" if its smallest dictionary contains more than 75,000 root entries AND it is not restricted to the translation of single-clause sentences AND it provides facilities for the user to extend the dictionary AND it supports a configuration in which more than one client can be networked with a single central server computer. Systems at this level are primarily meant for office use with networking capabilities.

There are some problems in the practical application of these definitions. The specification of particular components could mean the exclusion of systems from categories that appear appropriate, or the inclusion of systems in categories that seem inappropriate. For example, just because a system includes good facilities for users to create their own dictionaries does not mean it is more than a »basic« system, since the facilities included may be minimal or negligeable. The specification of dictionary sizes might imply some indication of 'quality' (i.e. the general public might believe that larger dictionaries must always produce 'better' results), and there is some clash with reality: systems for 'home use' may have larger dictionaries than some of those intended for 'company' use. In any case, the inclusion of specific sizes may become obsolete as systems and storage capacities in general become larger. In addition, definitions in terms of 'roots' are difficult to apply in practice because few vendors describe their dictionaries in these terms. It may be noted also that networking is no longer confined to 'advanced' systems.

For the »Compendium« it was decided to attempt simpler definitions oriented towards types of user and types of use, avoiding any suggestions of 'quality', expressed (as far as possible) in terms used by vendors when describing their own systems, and not specifying particular components or facilities, since in the »Compendium« the entries for particular systems include such information. The categories are intended to be general indicators of system types, defined in terms of targeted users.

The »Compendium« definitions for subcategories of 'fully' automatic systems (defined in 3.1 above) are:

MT system (home use): an automatic translation system designed by the producer for personal use by the general public, i.e. by persons not normally with any experience or training in translation.

MT system (for Internet/Web): system developed specifically for translating electronic documents on the Internet such as electronic mail, Web pages, chat discussions, etc.

MT system (professional use): system designed for use by professional translators, usually working independently or for translation agencies

MT system (client/server): system designed for company intranets to support a team of translators (often professionally trained)

It will be noted that the »Compendium« includes an additional category »(for Internet/Web)«. This was included because vendors make a distinction between systems (for »home use«) that are intended for use with word processing software on a stand-alone personal computer and systems (for »Internet« or »Web« use) that are designed specifically for on-line browsing.

4. Translation support tools

Both the IAMT Certification group and the »Compendium« identify the following types of support tools: electronic dictionaries, terminology management systems, translation memories, foreign language authoring systems, and (integrated) translator workstations (or workbenches). In addition, the »Compendium« has identified some further support tools.

In general, the definitions are closely comparable.

4.1. Electronic dictionaries

(a) IAMT Certification group:

A translation support tool is an »Electronic Dictionary« if it consists mainly of a bilingual or multilingual dictionary together with a dictionary lookup facility or mechanism.

(b) Compendium:

Electronic dictionary: bilingual or multilingual database of lexical entries (words or phrases) searchable individually or in combination, either for consultation or for insertion into human-produced translations.

The chief difference here is that the »Compendium« definition mentions the forms of entries (phrases as well as words) and the two basic types of use: 'traditional' dictionary consultation; and automatic insertion of translation equivalents. In addition, in the printed »Compendium« there is a note to emphasise the fact that some vendors sell dictionaries as 'translation systems':

(A dictionary might also be used to produce 'word-for-word' renditions, i.e. sequences of individually translated words in the syntactic order of the original texts.)

4.2. Terminology management systems

(a) IAMT Certification group:

A translation support tool is a »Terminology Management System« if it consists mainly of methods to help the user construct a multilingual terminology dictionary, together with a dictionary lookup facility implemented as a database (and not simply as a word-list), and includes multiple fields utilized according to standard terminology practice.

(b) Compendium:

Terminology management system: software for the creation, maintenance and searching of multilingual databases of terminology compiled for local (company or personal) use.

Terminology management was one of the first type of translation support tool, and is still one of the most popular among translators. However, increasingly, terminology management is combined with other translation tools, in particular translator workstations (see below 4.5).

4.3. Translation memory systems

(a) IAMT Certification group:

> A translation support tool is a »Translation Memory System« if it consists mainly of methods to help the user construct and use a collection of sets of previously translated text (ranging in length from phrases to whole texts) together with one or more of a text alignment facility, a text lookup facility, and a storage management facility.

(b) Compendium:

> Translation memory system: software (or component of system) for the creation, maintenance and searching of bilingual databases of previously translated texts.

The »IAMT Certification« definition is the more substantial of the two, attempting to explain how translation memories are (or can be) used. The »Compendium« definition is restricted to the core component. (It defines »Alignment tool« as a separate category, see 4.6 below.) The aim is to avoid any tendency to describe translation memory systems as if they were in themselves complete computer-aided translation systems, and it allows for the inclusion of translation memories as components of 'fully automatic' systems.

4.4. Foreign language authoring systems

Although there are still few systems on the market, this is a clearly definable category. The definitions differ only in emphasis:

(a) IAMT Certification group:

> A translation support tool is a »Foreign Language Authoring System« if it consists of lexicons, phrases, and even full text examples that assist the user to write documents such as business letters, contracts, etc., in a language in which they are not fluent.

(b) Compendium:

> Foreign language authoring system: software enabling composition of texts (e.g. business correspondence or electronic mail) in another language not necessarily from an original source text.

4.5. Translator workstation.

This is currently the most common type of translation support tool, intended for professional use, and primarily in large organizations – although increasingly there are 'workstations' being marketed for individual use.

(a) IAMT Certification group:

> A translation support tool is a »Translator Workstation« if it consists of several Translation Support Tools integrated into a single framework.

(b) Compendium:

> Translator workstation: integrated system for the use of professional translators, which combines (normally) multilingual word-processing, terminology management, translation memory, and (optional) automatic translation.

The main difference is the specification in the »Compendium« of the most common components, and in particular the inclusion of MT as an optional component. The latter feature is unfortunately one that serves to confuse the (theoretically) clear distinction between 'automatic translation' and 'computer-aided translation'.

4.6. Additional support tools

The »Compendium« identifies further support tools that are so far absent from the set defined by the IAMT Certification group.

As mentioned above, there is a definition of alignment, since there are commercial products designed specifically for this purpose:

> Alignment tool: software for the creation of bilingual text databases where sentences (or phrases) of source texts are linked to corresponding text segments of a target language.

Secondly (also because such specific products are marketed, although usually they are included in other software), there is:

> Pre-editing tool: software for the preparation of input texts, often including means for the control of input language, i.e. the reduction of ambiguities and the simplification of structures in order to facilitate automatic translation.

Finally, one of the most common categories of software is the variety of tools developed for the use of the localization industry:

> Localization support tool: system for the translation, terminological control, and publication of multilingual computer software documentation and programs

The inclusion of these support tools as a separate category in the »Compendium« was motivated by their large number and by the need to assist many users of the directory who would be looking for these aids.

At the EAMT Workshop it was pointed out that localization covers more that computer software and involves more than translation. A suggested wider definition is:

> Localization support tool: system for the cultural and linguistic adaptation of software and industrial products, including the translation, terminological control, and publication of multilingual documentation.

This definition (or a variant of it) will probably be included in future revisions of the »Compendium«

4.7. Internet translation services

Finally, the »Compendium« includes a definition for a type of Internet service that provides translations using computer software (as opposed to a service using human translators – which are also available on the Internet). This is not a 'support tool' (although it may be used for this purpose) but the provision of an automatic translation facility without requiring users to purchase any software. In effect it is, as far as level of function is concerned, a variant of the »MT system (for Internet/Web)« defined above (section 3.3). Since this type of service is being extended currently to mobile wireless devices (using the WAP protocol, etc.), the definition is fairly broad:

> MT service: translation service via Internet (or mobile telephone), using MT systems with or without human post-editing, and charging according to length and/or subject of texts. (Some on-line Internet services are free.)

5. Towards guidelines for evaluation of systems

Among the needs of potential purchasers must certainly be counted guidance in the assessment and evaluation of systems. While most large organizations can be expected to undertake their own evaluations and will have reasonably clear ideas of how to go about them, there is good evidence that the individual or small company users and potential users would welcome well-informed expert advice. Ideally, perhaps, they would like to see up-to-date evaluations of currently available systems formulated in terms easily accessible to those unfamiliar or ignorant of translation operations. It is, however, unlikely that IAMT will be able to undertake this function, although sub-contraction to another organization with greater experience in the testing of consumer goods might be an eventual option.

Before even this is possible, however, there is a need to establish more precisely what criteria should be adopted in evaluation. The criteria will ideally have to be applicable to the full range of translation software products (i.e. from fully automatic systems to specific support tools) and they will have to be easily (and cost-effectively) applied both in the evaluation of individual systems and in the comparison of systems. For this purpose, it is generally (but not unanimously) agreed that the IAMT could draw up a set of guidelines for companies and individuals to use when undertaking their own evaluations. The definitions of system types given in this paper might represent the first steps towards the establishment of such guidelines.

Traditionally evaluations have involved qualitative measures, such as faithfulness to the original message; intelligibility and comprehensibility; accurate rendition of terminology; and stylistic appropriateness (for the specific language and subject). Increasing use is made of measures for evaluating utility or usability: the savings of costs and time, in comparison with other systems or with wholly human translation; the ease of use (by experts, by novices); the level of intelligibility and/or accuracy, in accordance with the intended or expected application; speed and response times; training and setup costs; impact on an organization's overall translation throughput; compatibility with other systems (e.g. for authoring, publishing, terminology management, etc.) It has always to be stressed that systems suitable and cost-effective for one particular organization or individual may be quite unsuitable and uneconomic for another organization or individual.

Since the above definitions of systems and components have been formulated with particular regard to types of use and to the facilities and limitations of different systems types they could form the basis for guidelines for evaluations of utility. Such guidelines would be suggestions that users and potential purchasers of systems could apply in their own specific circumstances. They would not in themselves be methods of evaluation but only suggested evaluation criteria. An option for the near future would be for the IAMT Certification group (or some other working group) to compile a series of lists of criteria for different types of user. For example, a list of what companies might take into account when deciding on which systems to consider for evaluation, and when deciding how to evaluate; a set of factors that the 'professional' user (e.g. translator) might consider when assessing which system(s) to purchase; and guidance for the casual or occasional ('home') user about what a MT system can be expected to do and what it cannot do (at the current state of knowledge), and what types of translation support tools are available and how they may be used.

While the compilation of evaluation guidelines is a feasible near-future objective, the establishment of benchmarks must be seen as still, despite continued progress in evaluation methodology, a more distant goal. Benchmarks are, however, undoubtedly desirable for both vendors and consumers. Ideally, there should be benchmarks to measure the (comparative) performance of systems with respect to both translation quality and system usability. These could arise out of the application of agreed guidelines for evaluation. Whether the IAMT will ever be in a position to undertake its own benchmarking is an open question. More probable is that IAMT would co-operate with other associations in neighbouring fields in the formulation of benchmarks, and that it would sub-contract benchmarking tests to a reliable and trustworthy organization. At the least, however, I believe that one role of the IAMT could be to offer, through its members, advice to other organizations on setting up and applying agreed benchmarking tests. Such is the latent demand for benchmarking that if the IAMT does not undertake or sponsor authoritative tests of systems then it may well find that some other (perhaps less well qualified) organization may be doing it within a few years.

References

Hovy, E. (1998): 'Panel description: a seal of approval for MT systems'. In: Farwell, D., Gerber, L. and Hovy, E. eds. *Machine translation and the information soup. Third conference of the Association for Machine Translation in the Americas, AMTA'98, Langhorne, PA, USA, October 1998. Proceedings*, p.xi. Berlin: Springer, 1998.

Hutchins, J. (2000). *Compendium of translation software: commercial machine translation systems and computer-aided translation support tools*. [Geneva]: European Association for Machine Translation, on behalf of International Association for Machine Translation, April 2000.