T21N – Translation in Transition 2010-10

Nicole Keller (Heidelberg)

Integrating Multi-Word Terms in Terminology Management Systems: A Case Study

Editors:

Viktorija Bilić

Anja Holderbaum

Anne Kimmes

Joachim Kornelius

John Stewart

Christoph Stoll

Publisher:

Wissenschaftlicher Verlag Trier

T21N – Translation in Transition ISSN: 2191-1916 www.t21n.com

Nicole Keller (Heidelberg)

Integrating Multi-Word Terms in Terminology Management Systems: A Case Study

Abstract:

Terminology work in CAT-tool-integrated terminology systems has developed to become a key component of the translation process. The case study described in this article will focus on the reliability of automatic term recognition during the translation process in the editors of computer-aided translation tools. Using excerpts from medical texts as an example of technical texts, we tested the storage possibilities for, as well as term recognition and the final terminology check of, single-word and multi-word terms. As a "special case" of multi-word terms, collocations were also entered in terminology databases on the term level to check whether current systems offer an acceptable solution to this specific translation problem. The terminology system used is *crossTerm*, the terminology database of the corporate translation solution *Across*. It is an integrated system which offers a terminology database, automatic term recognition and terminology checking.

Die aktive Terminologiearbeit in Translation-Memory-Systemen hat sich inzwischen zu einem festen Bestandteil des Übersetzungsprozesses entwickelt. Dieser Artikel untersucht die Zuverlässigkeit der automatischen Terminologieerkennung im integrierten Editor eines Translation-Memory-Systems während des Übersetzungsprozesses. Als Beispiel werden Textausschnitte eines medizinischen Textes als Beispiel einer Fachtextsorte herangezogen. Dabei standen die Speichermöglichkeiten, die Terminologieerkennung und die abschließende Terminologieprüfung für die folgenden Kriterien im Mittelpunkt: Einworttermini und Mehrworttermini. Ein besonderer Fall im Bereich der Mehrworttermini stellen die Kollokationen dar, die ebenfalls als eigenständiger Eintrag angelegt wurden, um zu überprüfen, ob Translation-Memory-Systeme mit dieser Problemgröße sinnvoll umgehen können.
Exemplarisch wurde das Terminologiesystem crossTerm von Across verwendet, das Bestandteil der integrierten Übersetzungsumgebung von Across ist. Es beinhaltet eine Terminologiedatenbank, die automatische Terminologieerkennung und die Überprüfung der korrekten Verwendung der Terminologie.

Keywords:

terminology management system; multi-word term; single-word term; collocation; terminology recognition; terminology check.

Terminologieverwaltungssysteme; Mehrwortermini; Einworttermini; Kollokation; Terminologieerkennung; Terminologieprüfung.

Contents:

1 Introduction

In the last decade, terminology work in terminology management systems has developed into an essential part of the translation process using computer-aided translation (CAT) tools. One of the main features in the CAT-tool translation environment is the automatic recognition of terms. The importance of terminology work in achieving a consistent target language result has always been a major topic for translators. In the interest of efficient time management and quality assurance, translators should set up a database early in their career. This should be a database in which all terminology is gathered and stored from a range of sources and specialized fields. Such 'terminology management' (TM) basically

enables translators to compile a specialized encyclopedic dictionary on their PC. This can best be accomplished using computer-based management systems (Austermühl 2001: 102).

Consistent and correct terminology is one of the main problems in the translation process. For this reason, translation teaching has increasingly focused on the integration of terminology work in translation courses. Since terminology management systems have developed into an essential tool for translators, usually integrated in translation memory systems, there is also a wide range of courses focusing on the practical use of these systems.

The aim of this article is to determine the reliability of terminology databases used in the translation process, and to explain how quality in terms of terminological consistency can be guaranteed. In the case study, medical texts and terminology were used. Terminology recognition was tested for single-word terms, multi-word terms, and collocations as a special case of multi-word terms, because achieving their correct translation is one of the main problems in the translation process.

The terminology management system used for the demonstration is *crossTerm*, which is part of the corporate translation management tool *Across*. It is an integrated system which offers the available features (terminology database, automatic term recognition and terminology check) in all of the different product versions (Language Server and Personal Edition). Across Systems GmbH is headquartered in Karlsbad (Germany). The company is a spin-off of Nero AG and launched *Across* in 2003.

1.1 Historical Overview of Specialized Translation

The study of specialized fields (e.g. trade and agriculture) in translation dates back to the Middle Ages or even earlier. The first evidence of "technical" translations (medical, chemical, mathematical, and astrological) was documented in Ancient Mesopotamia. Focusing only on medical translation, however, leads us to one of the oldest volumes in medicine, the *Corpus Hippocratium*, that dates back to the 5th century BCE (Montalt and González Davies 2007: 15).

Generally speaking, ever since humans have communicated with each other they have been confronted with translations. But the study of technical language, also known as specialized

language or language for special purposes (LSP), only became a part of modern linguistic studies in the 1930s and 1940s. Since then, the problem of specialized terminology has been one major focus of these studies. When computer-aided terminology databases were first introduced in the 1960s, this represented a completely new field within the study of language for special purposes (Lee-Mittelbach 2008: 5-6).

Today, terminology work is generally directly integrated into the translation process, especially in the field of technical texts that are translated using computer-aided translation tools such as *Across, SDL Trados or Transit*. Terminology databases are in most cases default components of standard CAT-tool packages and automatically form part of the translation environments. Nevertheless, users of these systems must realize that terminology work does not take place automatically, but has to be done manually by each user. This means that, although there are features available such as terminology extraction and text analysis functionalities that assist the translators in finding terminology from a pool of words, selection and translation of the relevant terms must still be performed by the users themselves.

1.2 The Challenge of Specialized Languages

During the last few decades, specialized language has become increasingly important, since knowledge in a broad spectrum of fields in science and technology has also increased dramatically. As a result, communication between specialists and laypeople has become more and more difficult because numerous new terms have been created and defined as new phenomena and products have been discovered, invented, and named. These problems affect intralingual and interlingual communication alike. Communication between different languages is comparatively more difficult but has also increased in importance because of the increased need for translation in a global economy. Translation of technical texts is only possible if a translator is knowledgeable about the terminology in the special field, and this goal can most often only be attained through thorough preparatory terminology work (Arntz, Picht and Mayer 2004: 1).

Depending on the existing knowledge of the translator, translating specialized texts from one language into another can be very time-consuming and is accompanied by further translation-related problems:

- understanding the text content in the source language
- writing a target text according to the standards used in the target language, which can differ considerably from those of the source language
- solving translation problems
- detecting and defining terminology in source and target language.

Using incorrect terminology or phrases that are uncommon for this particular field of technical language will inevitably lead to non-acceptance of the produced target text and expose the translator as a non-specialist.

Considering a medical text, for example, Marla O'Neill (1998), a medical doctor with a strong linguistic background, discussed whether the medically knowledgeable linguist or the linguistically-knowledgeable medical professional is the better medical translator. She created a survey that was sent to medical translators asking them about their professional backgrounds and their everyday work. Her conclusion was that medical translations can be produced by both medical professionals and medically-knowledgeable linguists. Nevertheless, she concludes that the ideal situation would involve the editing by linguists of translation texts produced by medical professionals and vice versa. The key competencies required by both are the specialized knowledge of the field, an "ear for style," and a sense of writing style and skills.

Style and writing skills are difficult to learn and are probably based on a long-term learning process. Computer-aided translation tools are limited in their abilities to detect and correct grammatical, orthographical and stylistic errors. There are special programs that check texts on the basis of controlled language patterns, which are already used in the field of technical documentation, but these programs are usually restricted to user-defined criteria and are not generally applicable to a wide range of specialized fields.

Terminology databases, however, can offer support to the medical translator in using terminology consistently in the target language and can automatically check the correct use of terminology in an existing translation. The questions remain: Is it always possible to store and use terminology? How reliable are terminology management systems? And can they be trusted blindly?

2 Basic Principles of Terminology Management

2.1 Concept

Terminology work in computer-based terminology systems is concept-oriented, which means that only one meaning is entered per entry. For example, the term "shift" has two different meanings (concepts): *change* and *period of work*. In a general bilingual dictionary, which is usually term-oriented, the two meanings would be explained below the main entry of the term separated by *1.* and *2.* to make clear that two separate meanings exist. In a terminology database, the two meanings are separate entries.

Term-oriented (usually bilingual):

EN: shift
1. EN: change DE: Veränderung
2. EN: period of work DE: Schicht
=> Language direction cannot be reversed.

Concept-oriented (can be multilingual):

EN: shift DE: Veränderung
EN: change

EN: shift DE: Schicht
EN: period of work
=> Language direction can be reversed.

Using this principle, it is possible to build up a multilingual database with as many languages as necessary, and this is the main advantage of a concept-oriented approach. A translator

gains the ability to create translations between all language combinations represented and stored in the terminology database.

The semiotic triangle first proposed by the linguists Ogden and Richards in 1923 clearly illustrates the concept notion.

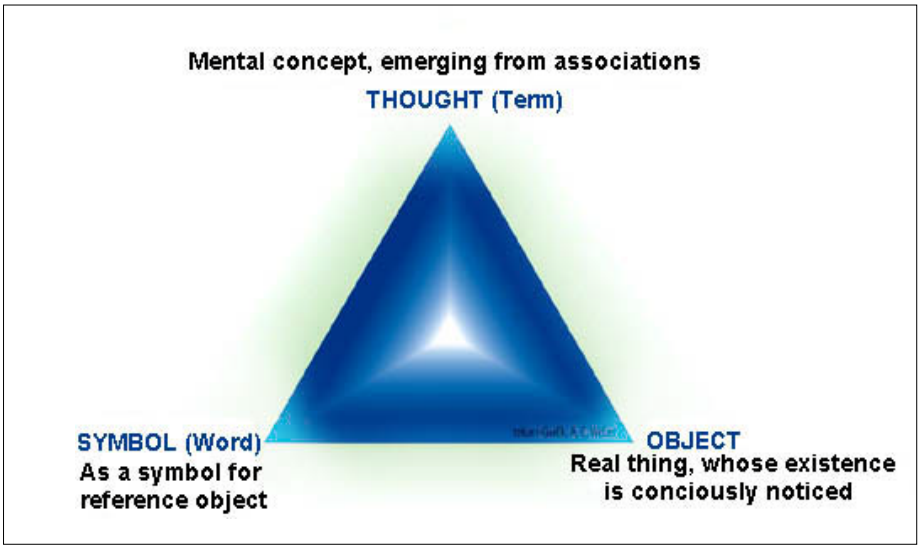


Fig. 1: The Semiotic Triangle (Sieber and Chen 2007)

The various elements of the semiotic triangle can be defined as follows: An **object**: is "any part of the perceivable or conceivable world" (ISO 1087). Objects can be material (e.g. a certain building, the Sydney Opera House) or immaterial (speed, pain) (DIN 2330). A **concept** represents "a unit of thought constituted through abstraction on the basis of properties common to a set of objects. The semantic content of a concept can be re-expressed by a combination of other and different concepts, which may vary from one language or culture to another" (ISO 5963). Concepts are not connected to any specific language (DIN 2330). **Term**, finally, refers to a "designation of a defined concept in a special language by a linguistic expression" (ISO 1087).

2.2 Translation-Oriented Entry Structure

The following description of a translation-oriented entry structure for terminology databases is based on the ISO Standard 12616, which is exemplary for the majority of entry structure

models. A comprehensive overview of different entry models is given in Keller 2006 (chapter 1).

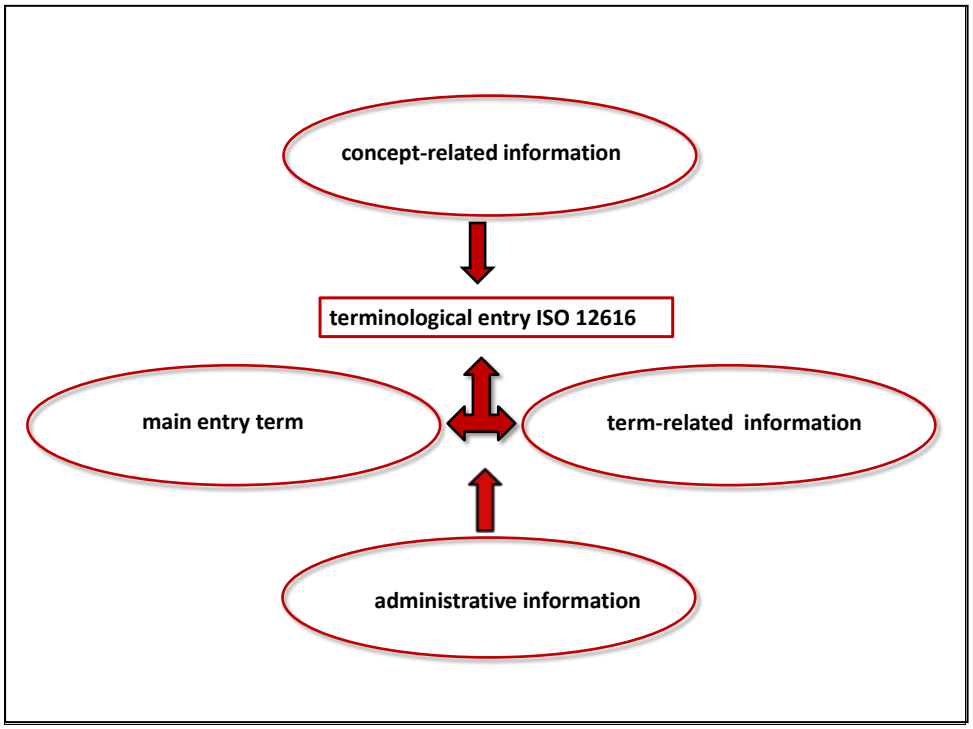


Fig. 2: Terminological Entry in Accordance with ISO 12616

The most important data category described in the ISO standard is the *main entry term* (short form: *term*). Every designation that can be used as a synonym describing one concept is entered in this category, e.g. abbreviations, variants, formulas, symbols, etc. In a terminological database, this is also the main data category to be filled in because it is the index field. Whatever system is used, it is essential to indicate at least one main entry term in each terminological entry (ISO 12616). In terminology systems, the main entry term is usually part of the term-related category since corresponding fields define the term itself.

The *term-related data category* contains all information that is descriptive of the term itself. This includes grammatical information (e.g. grammatical gender), usage information (e.g. temporal qualifier, when the usage has changed over the years), status information (e.g.

term status, 'preferred' vs. 'not to be used'), and equivalence information (e.g. directionality, unidirectional in case of specific country connotations).

The *concept-related data category* represents all types of explanatory material used to elucidate a concept. This includes information on domains and subdomains specifying the classification of the concept within a subject framework. Furthermore, descriptive elements are a major part of this category, providing definitions or explanations of a concept.

The *administrative data category* consists of information on how up-to-date the information is and makes it possible to keep track of changes. Most of the information in this category is automatically added by the system, since the date and name of the responsible user are taken from the system settings.

Detailed information on other data categories for terminological entry models can be found in ISO 12620, which lists possible data fields plus explanations for terminology work in general.

Recommended guidelines for accurately filling in these data categories are outlined and described in the introduction to terminology work by Arntz, Picht, and Mayer (2004).

3 Terminology Work and Term Recognition: *crossTerm* as an Example

This section describes the principles of practical terminology work with terminology databases using the example of *crossTerm*, the terminology database from *Across*. In the following case study, the focus will be on the storage of single-word terms and multi-word terms.

In accordance with DIN 2330, single-word terms consist of one word and include compound words and hyphenated compound words as well.

Multi-word terms consist of at least two words that are separated by a space character (DIN 2330).

crossTerm consists of four separate windows reflecting the four principles: main entry term, concept-related, term-related, and administrative data categories as described in the ISO Standard 12616 (Figure 2). In addition, *crossTerm* also integrates a flexible search pane.

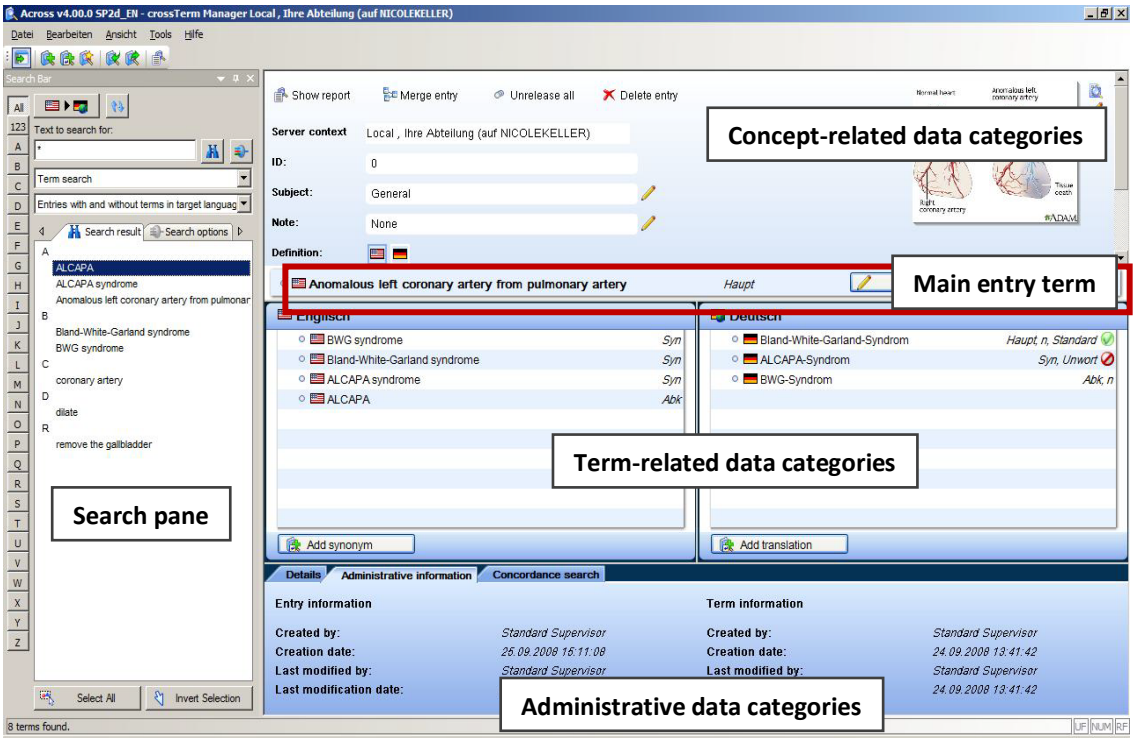


Fig. 3: Description of the Terminology Database *crossTerm*

3.1 Entering Terminology

In order to fill the database with information, *Across* offers a number of different approaches.

a) Workflow: Terminology work

When a document is imported into *Across* during the creation of a new project, the project manager can define a workflow that always includes translation of the document, but may also include proofreading of terminology work and the translation. Terminology work is divided into two separate tasks: term extraction and term translation. During the term extraction process, a list of term candidates is extracted from the source text and presented

to the user for evaluation. After marking all the terms to be translated, the selected terms are handed over for term translation. A translator translates the suggested terms using the source text and the existing information from the *crossTank* translation memory and *crossTerm*. After finishing this task, the new terminological entries are stored in *crossTerm* and can be used in future translations.

b) Quick copy from the translation environment *crossDesk*:

During the translation process, the translator can copy source terms and their translations directly to *crossTerm* by simply highlighting them in the source and target text. *Across* immediately switches to *crossTerm* with both terms already entered. In the next step, the user has the option to add additional information, such as grammar information, usage notes, or context examples. After saving the new terms, they are immediately shown in the *crossTerm* window integrated in the translation editor *crossDesk* and can be used in the translation process.

c) *crossTerm Manager*

Within the administration component of the terminology database, the *crossTerm Manager*, the user can import existing terms from other terminology databases or from *Excel* lists. A simple mapping process allows for the efficient integration of existing external information.

It is also possible to simply open the edit mode in *Across* and enter terminology from scratch. The option *Quick add terms* is a slightly restricted but a quicker way of entering terms directly into the *crossTerm Manager*.

d) *crossTerm Web*

The Web interface for accessing the information in *crossTerm* without an existing *Across* installation allows the user to enter and edit terms and access the information via the Web. In this way, users from all over the world can collaborate to increase terminology related to a company, and terminology work can be organized in a decentralized way.

3.2 Terminology Recognition

Once terminology is entered into the system, it is automatically recognized in the *crossDesk* translation environment. In this context, the following questions are important: How does automatic term recognition work, and is it reliable? Are there any differences between recognizing single-word and multi-word terms? Is there also an efficient way of integrating collocations as terms?

For the following case study, a medical English text about a condition called ALCAPA (anomalous left coronary artery from pulmonary artery) is used as the source text for a translation into German. Before the translation is produced, three terminological entries were defined and added to the terminology database in the language pair English-German. The German translations are also marked with the additional information *recommended*, *synonym*, and *do not use*. The main focus was to have single-word terms that can appear in different word forms (plural for nouns and past tense for verbs) and further to have multi-word terms with a different number of words to see whether the system will recognize all of them. Furthermore, slight changes were made to the multi-word terms that are explained in detail in the examples below.

English	**German**
anomalous left coronary artery from pulmonary artery ALCAPA ALCAPA syndrome Bland-White-Garland syndrome BWG syndrome	Bland-White-Garland-Syndrom (recommended) BWG-Syndrom (synonym) ALCAPA-Syndrom (do not use)
Artery	Arterie
Dilate	erweitern

Table 1: Terminological Entries Stored in *crossTerm*

Figure 4 shows an active paragraph in *crossDesk* displaying the automatic term recognition. The terms from the database are detected in the source text and marked with a red over-line. On the right-hand side, the *crossTerm* window displays the terms that appear in the

active sentence and their suggested translations into German. Looking at the red over-line, it becomes clear that the terms mentioned above are found in the text and that the past tense form of the verb *dilate* and the plural form of *artery* are also marked. This is because *Across* supports the stemming of words when searching for terms in *crossTerm*, and therefore finds inflected forms of words as well.

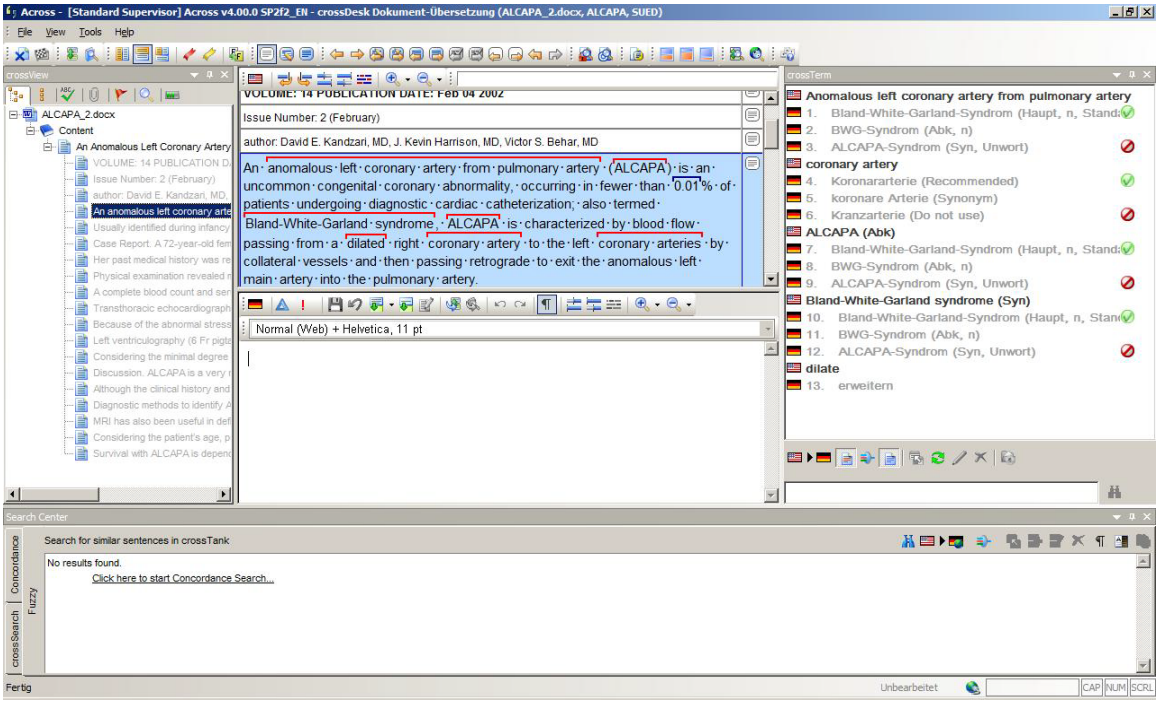


Fig. 4: Translation Environment *crossDesk* in Across

In Figure 5, some parameters of the source text were changed to see how flexible and reliable the multi-word term recognition is. The relevant text passages are marked in red to underline the differences compared to the original text.

- The term *anomalous left coronary artery from pulmonary artery* was set in plural form to check whether the stemming also works for multi-word terms.
- In the second part of the sentence, the word *anomalous* is omitted to check whether parts of a multi-word term are also recognized.
- The "spelling" of *Bland-White-Garland syndrome* in the source text was changed and is written without any hyphens.

- Finally the word order of *ALCAPA syndrome* was changed into *syndrome ALCAPA* to check whether switched words are still found.

The results indicate that only the plural of multi-word terms is still recognized. Different writing variants or even misspellings, omission of words, and the switching of words do not provide any matches in the terminology window.

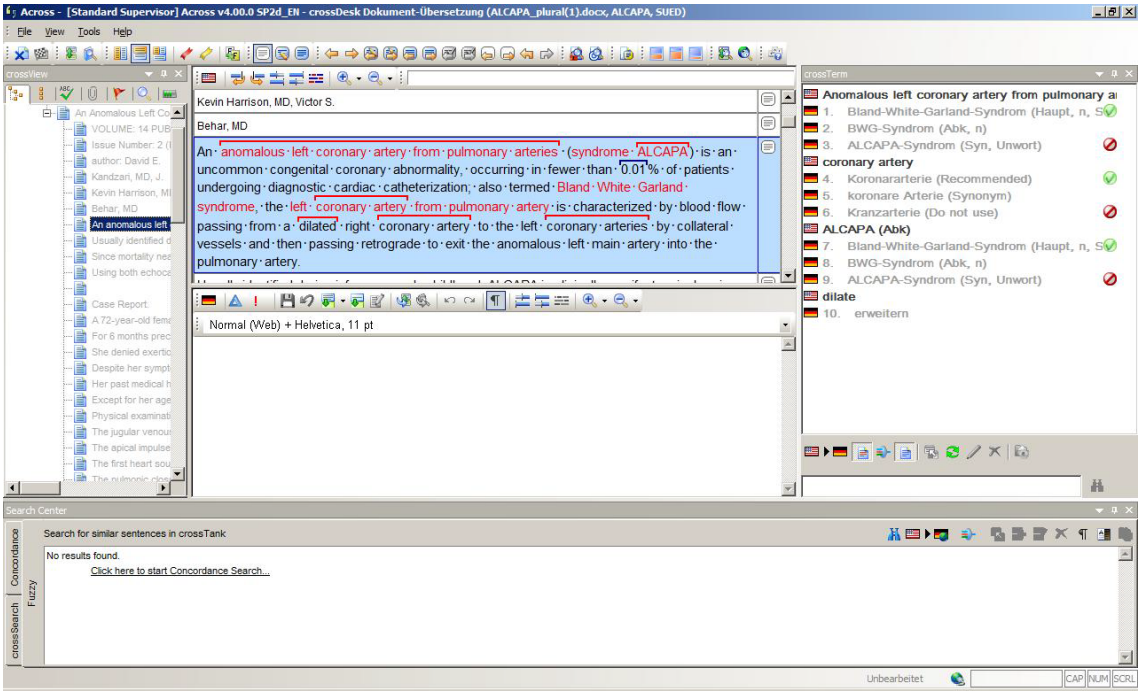


Fig. 5: *crossDesk* With Slight Parameter Changes

In this context, it is important to keep in mind that the prepared text passages were used for test purposes, and multi-word terms are usually used in their natural word order. It is not very likely that the word order would be switched or that parts of a multi-word term would be left out because they are seen as a fixed unit that is meant to appear in the same format in which they are stored in the terminology database. These examples demonstrate that there can be situations when stored terminology appears in the text and is not automatically shown to the translator.

However, terms can be written in different ways, especially when it comes to hyphens; many people have their own spelling. Finally, in real life situations, texts are often imperfect documents containing spelling errors.

3.3 Summary

Terminology databases such as *crossTerm* support the modern translation process and greatly empower the translator. Both single-word and multi-word term entries produce reliable automatic term recognition in the majority of cases. Nevertheless, as evidence in the examples shows, the individual translator remains responsible for terminology and must be prepared to manually look up existing terms that do not appear automatically.

The integration of further functionalities, such as a more detailed part-of-speech recognition or the option "find nearby" or fuzzy search could further improve the already high reliability of terminology databases.

3.4 Collocations as a Special Case

Within the project "Collocation Research in the Science of Translation" collocational occurrences in technical and medical texts (to take two examples) have been the focus of research at the Seminar for Translation and Interpreting at the University of Heidelberg. Research on the integration of collocations as terms in terminology databases developed from these projects.

A collocation refers to a combination of lexical terms that tend to occur together in common language use. Such combinations are semi-fixed in nature, meaning that they are not always interchangeable with other lexical terms. Doing so would result in an unidiomatic-sounding phrase, which would be automatically detected by a native speaker of the language. For example, an English native speaker would recognize that the word combinations "to commit a crime," "to express an opinion," and "gnawing hunger" are valid collocations and "to take a crime," "to talk an opinion," and "itching hunger" are erroneous (Lee-Mittelbach 2008: 21).

The translator must learn these word combinations—a daunting challenge involving an enormous amount of information belonging to a specialized field of knowledge.

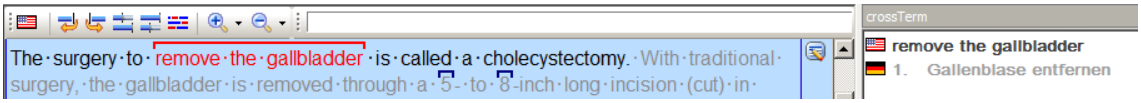
One solution for this translation challenge would be to integrate collocations in terminology databases. Various possibilities are feasible, including creating a separate data category for collocations or adding collocations as term attributes within the term-related category. Connecting a terminology database to a collocation dictionary is also a reasonable option. Nevertheless, all of these share the disadvantage that during the translation process, the collocation is not recognized or automatically displayed.

The most practicable and useful integration of collocations in terminology databases seems to lie in creating an entry for each collocation. Since each collocation has its own definition, and the concept-based principle of terminology databases requires separate entries for different meanings, this entry system offers a feasible solution. Within terminology work, collocations are considered equivalent to multi-word terms, and therefore could be automatically detected during the translation process.

The question arises: Can collocations be completely treated as multi-word terms or are there any additional problems? The definition of multi-word terms indicates that all words that belong to a multi-word term form a fixed unit and are not very likely to be changed. Collocations, however, are semi-fixed word combinations and can appear in different ways.

For the following example in *crossDesk*, the collocation *remove the gallbladder* (German: *Gallenblase entfernen*) was entered into the terminology database. The source text shows several different forms of the collocation that are highlighted in red:

1. *Remove the gallbladder* (natural order as it appears in the database).

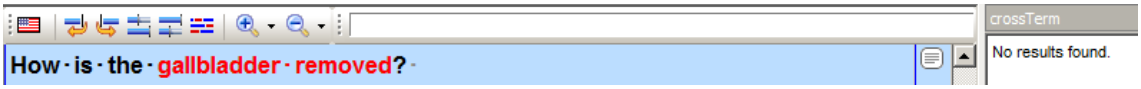


The collocation is recognized and the translation is displayed in the *crossTerm* window.

2. *Gallbladder removal* (change of collocation type: verb + noun changes into noun + noun).

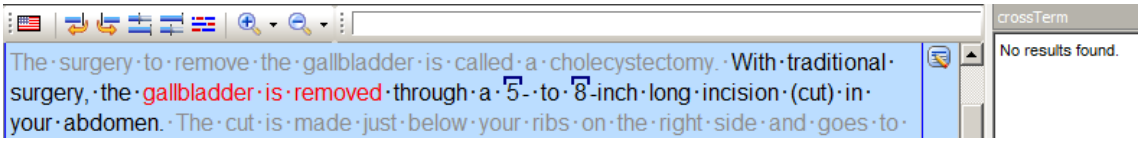
The change of the collocation type is not detected by the system since there is no part-of-speech recognition.

3. *How is the gallbladder removed?* (change of word order determined by the sentence structure: question).



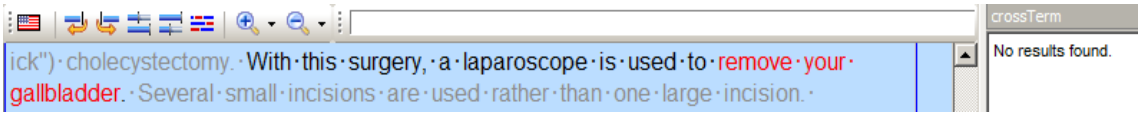
The change of word order is more likely to appear with collocations than with proper multi-word terms since the collocation type noun + verb, for example, has to match the sentence grammar.

4. *The gallbladder is removed* (passive voice).



Passive voice is equivalent to the change of word order and is therefore not recognized either. This example also covers a present perfect or past tense example such as "the gallbladder has been/ was removed."

5. *Remove your gallbladder* (slight changes of the collocation itself).



Slight changes are also more likely to appear in collocations than in multi-word terms since the construction itself is not as fixed.

3.5 Collocations: State of the Art

Although collocations can be integrated into terminology databases, their automatic recognition and subsequent translation remains inadequate. Term recognition only functions when the collocation appears in its natural order. Since the nature of collocations involves their appearance in various forms, especially when one part of the collocation type is a verb, their automatic recognition during the translation process is insufficient. This represents the major difference between collocations and multi-word terms.

If translators decide to store collocations in terminology databases they have to keep in mind that the usage is limited to manual look-up.

4 Quality Assurance: *crossTerm* Check

The *Across* system provides an integrated terminology usage check feature to assure the quality of the target text. The system checks for terms in the source language that are stored in the terminology database and scans the target text for the correct translation as suggested in the terminological entry.

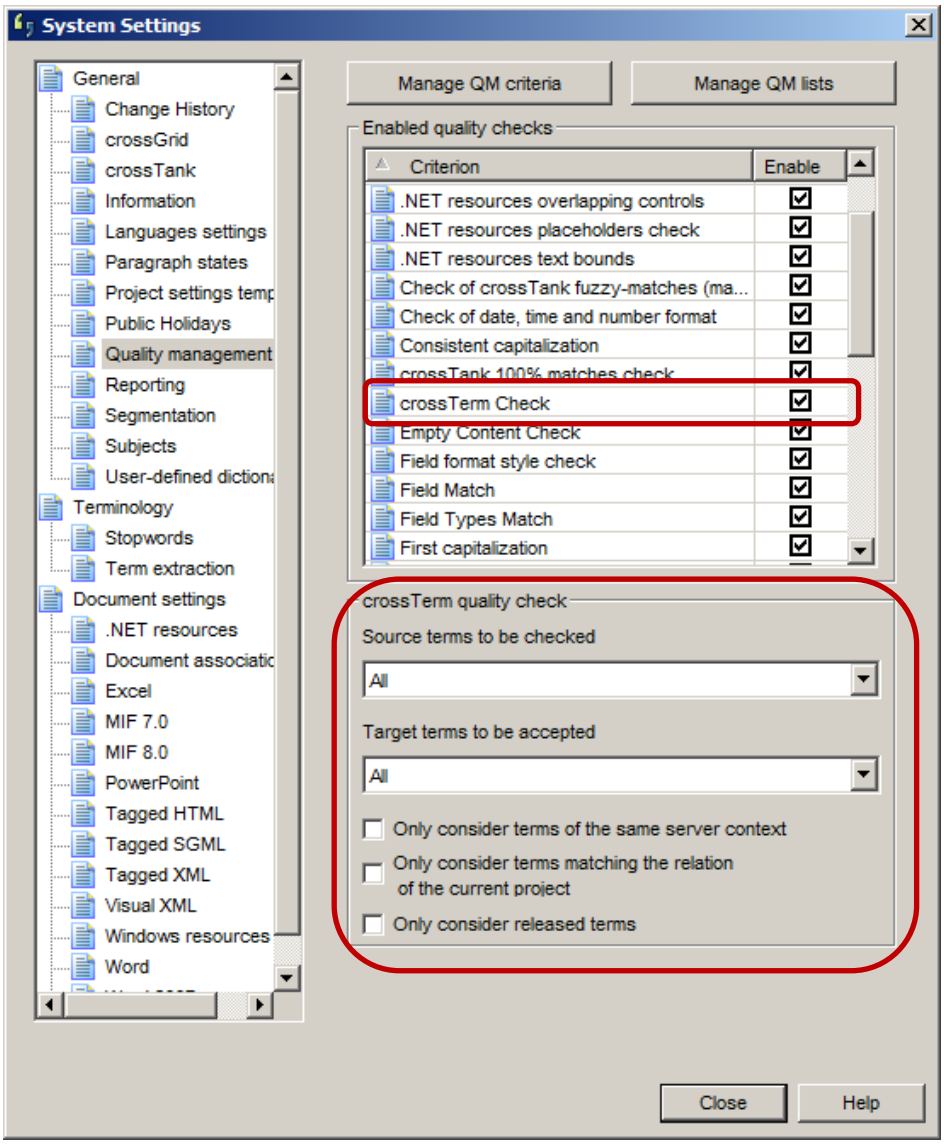


Fig. 6: The Quality Management Tab in *Across's* System Settings

The *crossTerm* quality check can be defined more precisely. As explained in section 3.2, terms can be evaluated and indicated by the flag *recommended*, *synonym*, and *do not use*. The terms can be checked in accordance with their usage flags. There are three separate options for the terms that are either in the source or the target text. First, both texts can be checked for all terms, regardless of their flags. In this case, *do not use* terms are only seen as *synonyms*. Second, only terms that are marked as *recommended* are checked in the source

text, which means that terms that are either marked as *do not use* or *synonym* are not checked at all. For the target text the effect is slightly different because the text is checked for terms that are marked as *recommended*, and all terms that have the flag *synonym* or *do not use* are marked as incorrect. Third, if terms are marked as *do not use* in the source text, the system will only check these terms and their translations. If the option *all but do not use* is activated for the target text, the check will also accept terms that are marked *synonym*.

Since *Across* stores all entries in the same database, and not all of the terms in a translator's database might apply to the current translation, it can also define a certain group of terms that match at least one attribute, like *relation* (usually the name of the customer you work for), *server context* (description of the data source; for example, when translators work for several customers by connecting to their servers and receiving tasks they will have data from different server contexts stored in the same database), or *released status* (refers to the status of a term; after evaluation by a terminologist the terms can be assigned the status 'released'). This represents another option for improving the quality of the terminology check.

In the following steps, the terminology check will be tested for three separate occurrences of terminology in medical texts: single-word terms, multi-word terms, and collocations.

4.1 Terminology Check for Single-Word Terms

In the following example, the sentence "Except for her age and post-menopausal status, she had no significant risk factors for coronary artery disease" is to be translated into German. As we can see in Figure 7, *Across* detects the term *artery*, offering the translation *Arterie*. As a medical translator can see, however, the adequate solution in this example sentence would be the multi-word term *coronary artery*, which has not been entered into the terminology database. In this case, the translation into German would be *Koronararterie*, a commonly-used compound noun in German medical texts. Depending on the spelling of the translated word, the terminology check will detect an error. If the translation is *koronare Arterie* (which is unusual in German but the only way to have *Arterie* as a separate word in the German translation) or *Koronar-Arterie* (which is an unusual spelling but also offers

Arterie as an independent word) the system will not detect an error. But if the translation *Koronararterie* is used in the target text, *Across* will detect an error since *Arterie* is only part of a compound noun. This indicates that the check involves the application of formal rules rather than linguistic ones, and that the results of the check vary greatly depending on the language combination.

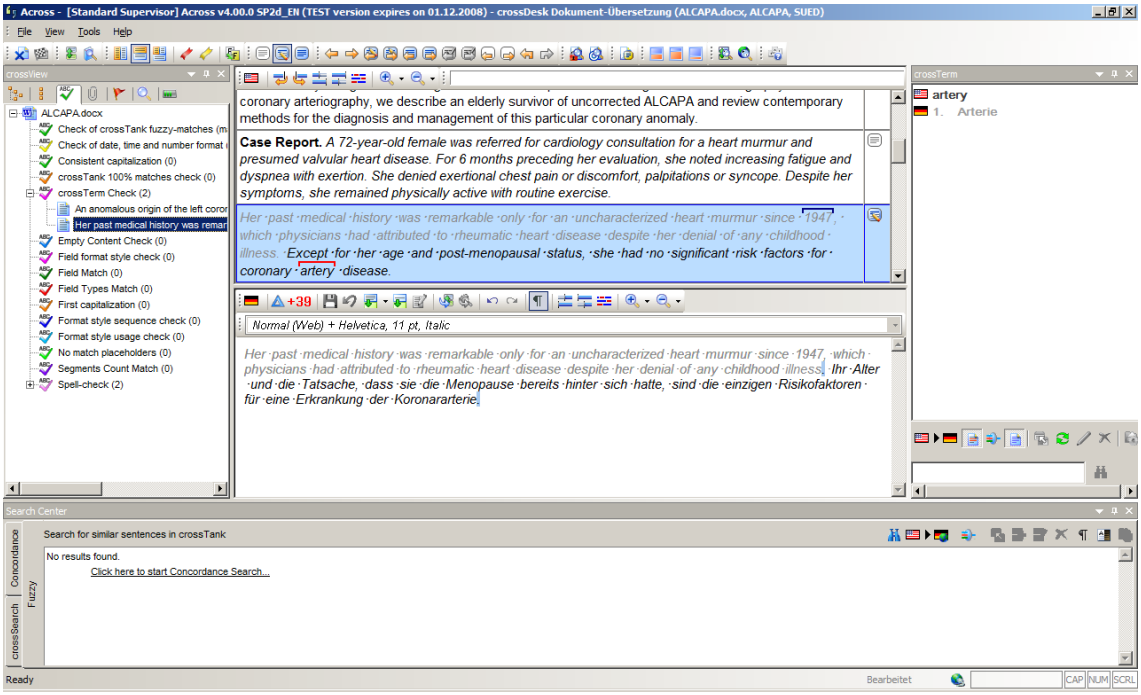


Fig. 7: *crossTerm* Check of Single-Word Terms

In general, the check for single-word terms that are proper single words (meaning they are not part of a multi-word term such as "coronary artery" that has not been entered into the terminology database) is a good and reliable feature in *crossTerm*.

In section 3.2, we demonstrated that plural forms of nouns (*arteries*) and inflected forms of verbs (*dilated*) are also found by the automatic terminology recognition. But the term check in the target text does not necessarily detect the varied forms as well. *Arterien* was also labeled as correct, but the inflected form of *erweitern* in the German phrase *erweiterten Koronararterie* was considered not correct. This is another indication that linguistic rules are not applied.

To verify whether terminology recognition and terminology check in fact work in the same way, a German text containing the phrase *erweiterte Koronararterie* was imported as a source text. The result was that neither *erweitern* nor *Arterie* was displayed in the *crossTerm* window.

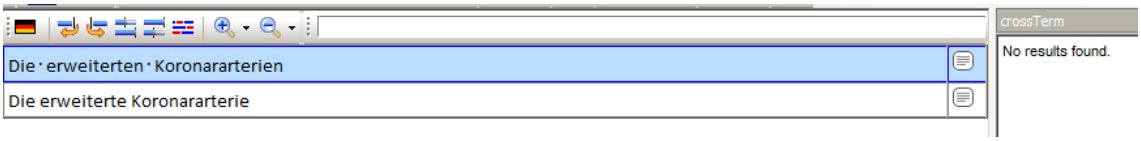


Fig. 8: No Results Found for *Die erweiterten Koronarterien*

4.2 Terminology Check for Multi-Word Terms

In the example for the multi-word term, the term *anomalous left coronary artery from pulmonary artery* was used in the source text with the German translation *Bland-White-Garland-Syndrom*. The term was recognized and the translation checked as correct.

In a second translation, the spelling of the German translation was slightly changed into *Bland-White-Garland Syndrom*, omitting the last hyphen, which is incorrect though not unusual. In this case, the term check does find an error. One example of the desirability of spelling consistency in multi-word terms is *email* or *e-mail*. Though it appears in various spellings, when it comes to terminology checking, a stricter control of the word's spelling assures consistency in writing of technical terms.

In another test, the German translation was changed slightly into *Syndrom Bland-White-Garland*, and the next time *Syndrom* was omitted. Neither term was recognized by the terminology check since the terminology recognition did not work for those cases either. But again, during the terminology check this result is more acceptable than for the term recognition feature, as using the exact word combination stored in the terminology database is crucial for assuring consistency.

The check was also carried out for a plural form in the target language, which would generally not be considered an error. The original sentence contains the term *coronary arteries* and is recognized by the system. The translation contains the term *koronare Arterien* and is not detected as an error. In this case a two-word term was intentionally chosen as the

translation and not the more common translation *Koronararterien* since this is a single-word term.

Provided that a multi-word term is correctly recognized in the source language, the check for translations is a reliable tool for translators, since correct spelling in the target language is crucial in a text production task that requires correct terminology. The main difference to the source text is that during text production in the source language, spelling and terminological mistakes frequently happen since in most cases no controlled language or terminology control is used during the initial process. During the translation process, it is essential that the required terminology is used.

4.3 Terminology Check for Collocations

In section 3.3 we indicated that in many cases collocations are not detected by the term recognition functionality because of the fact that collocations often appear separated, reversed or in a slightly modified form, especially where verb + noun combinations are concerned.

In this section, only the translation of detected collocations will be checked because the check for the other cases cannot work properly without a reference term. The collocation to be checked is *remove the gallbladder* with the German translation *Gallenblase entfernen*.

The trial sentence is: "The surgery to remove the gallbladder is called a cholecystectomy." To test the "collocation check" several translations are provided with typical phenomena that are likely to appear during translations. Some translations have slight meaning variations to demonstrate the differences.

First translation: *Cholezystekomie bedeutet die Gallenblase entfernen*. The collocation appears as in the terminology database and no error is reported.

Second translation: *Cholezystekomie bedeutet die Gallenblase zu entfernen*. The collocation appears as in the terminology database, but since one additional word was inserted ("zu"), the system reports an error.

Third translation: *Cholezystekomie bedeutet, dass die Gallenblase entfernt wird*. Since the collocation is in the passive voice, the check reports an error.

Fourth translation: *Das Entfernen der Gallenblase nennt man Cholezystekomie*. Nominalization of the verb and changing of the word order results in an error report.

During the collocation recognition process, it became clear that the lack of linguistic rules does not lead to a satisfactory result. The same applies to translated collocations, especially with languages such as German, which uses inflected forms and varying word order depending on the sentence structure. It appears that the current storage capability in terminology databases is only appropriate for manual research, but not reliable for auto-detection.

4.4 Terminology Check for "Recommended" and "Do-not-use" Terms

Occasionally, translators use various terms for one concept if there is no corporate language defined. This occurs especially when translating into German, where it is considered to be bad style to repeat the same word too often. Over the years, a company can accumulate a lot of different terms for one and the same thing. Finally, a company collects all the different terms and stores them in a database to define an official corporate language.

In these cases, as explained above, terms in *Across* can be marked as *recommended* and *do not use*, designated by a flag which includes a special attribute plus a visible icon with the term (see Figure 10).

The usage of *Kranzarterie* and the option *Target terms to be accepted: Recommended only* in the system settings (see Figure 6) detects all terms in the translation that are used and not marked as *recommended*. If this option is not activated, *crossTerm* would not detect an error since *do not use* terms are also synonyms and therefore correct.

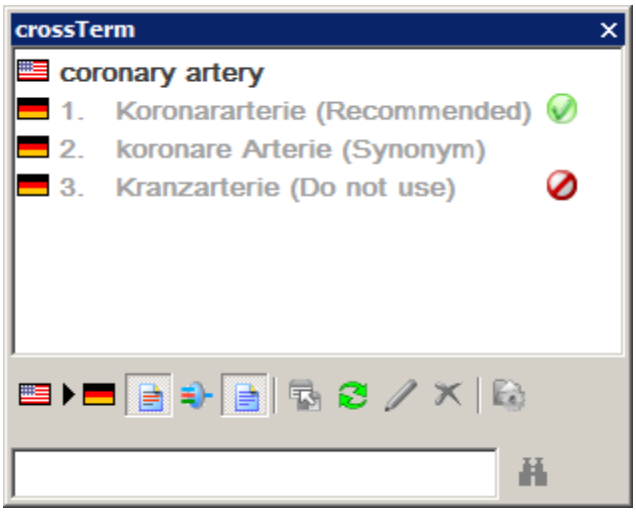


Fig. 9: Recommended and Do not use Flag in *crossTerm*

5 Conclusion

The analysis in sections 3 and 4 showed that today's terminology management systems are useful and effective tools in the translation process. Automatic term recognition supports the translator, and terminology checks ensure consistent term usage in the target text. These functions work efficiently for single-word terms and when terms appear in their basic or slightly altered forms, e.g. in the plural.

However, it was also evident that there are situations involving single-word terms containing slight changes, such as irregular inflections or compound nouns, when the system fails to provide the user with a result. Multi-word terms are often not recognized at all by the system because of varying spelling, hyphens, changed word order, additions or omissions, or other minor changes, such as changing an article (the) into a possessive pronoun (his). Depending on the source and target language, these problems can appear more or less frequently.

Essentially, these problems occur because no linguistic rules are applied to the process of terminology recognition and control. Since these systems usually support numerous languages, it is clear that implementing linguistic rules for all languages and all language combinations represented is currently basically impossible. Yet, although not all world languages have yet been covered by the field of computer linguistics and machine

translation, comprehensive systems have been created for frequently used languages, such as English and German. One might ask if the established rules and algorithms can also be implemented for remaining languages.

At the moment, translators should be aware that the terminology database generally contains more terms than are immediately evident. In order to access and use the full range of terminology available, a manual search is occasionally fruitful, particularly in those cases when the translator notices that terms are not appearing in their basic forms. This also applies to the terminology check, which only works for those terms that are recognized by the system, and is therefore more restricted.

The case study showed the importance of integrating terminology work and the practical use of terminology management systems into translator training since the everyday use of computer-aided translation tools and the integrated terminology component is an essential competence translators have to learn in this day and age. The intelligent and active use of the new technologies will definitely increase the quality and consistency of the translation. Nevertheless, these systems are designed to support (aid) translators, not to replace them. Hence, a translator should never place complete trust in CAT tools but should benefit from their wise and judicious use.

6 References

Across Systems GmbH (2008). *User Manual 'Across at a Glance.'* Karlsbad: Across.

Arntz, Reiner, Heribert Picht, and Felix Mayer (2004). *Einführung in die Terminologiearbeit*. Hildesheim: Olms.

Austermühl, Frank (2001). *Electronic Tools for Translators*. Manchester: St. Jerome.

DIN 2330 (1993). *Begriffe und Benennungen. Allgemeine Grundsätze*. Berlin: Beuth.

ISO 12620 (1999). *Computer Applications in Terminology – Data Categories*. Genf: ISO.

ISO 12616 (2002). *Translation-oriented Terminography*. Genf: ISO.

Keller, Nicole (2006). *Neue Wege in der Hilfsmittelkunde der Übersetzungswissenschaft: Zur Herleitung webbasierter Terminologiedatenbanken im Kontext von CAT-Systemen*. Trier: WVT Wissenschaftlicher Verlag Trier.

Lee-Mittelbach, Karen (2008). *Dissecting Medical Case Studies – A Linguistic Analysis*. Heidelberg: WVT Wissenschaftlicher Verlag Trier.

Montalt, Vincent and Maria González Davies (2007). *Medical Translation Step by Step*. Manchester: St. Jerome.

Ogden, Charles Kay and Ivor Armstrong Richards (1923/1999). *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. London: Continuum Publishing.

O'Neill, Marla (1998). "Who Makes a Better Medical Translator: The Medically Knowledgeable Linguist or the Linguistically Knowledgeable Medical Professional? A Physician's Perspective." Henry Fischbach (ed.) *Translation and Medicine*. Amsterdam and Philadelphia: John Benjamins. 69-80.

Sieber, Tanja and Willy Chen (2007). "Intelligent Products Find Their Errors Themselves." *Technische Kommunikation* 29.6. 28. http://www.tekom.de/index_neu.jsp?url=/servlet/ControllerGUI?action=voll&id=2268.

T21N - Translation in Transition

T21N offers a cutting-edge electronic publishing venue, created by experts for both young talent and established researchers from the worlds of translation and interpreting.

T21N provides a stage for emerging ideas and new academic talent to present their ideas in a digital reading site, where speed and ease meet enjoyment.

T21N is exclusively published online at http://www.t21n.com.

Articles in compliance with our style sheet may be submitted at any time and will be published at short notice.

T21N editors research and teach at the Institute of Translation and Interpreting at the University of Heidelberg in Germany.

Editors:
Dipl.-Übers. Viktorija Bilić, Dr. Anja Holderbaum,
Dr. Anne Kimmes, Prof. Dr. Joachim Kornelius,
Dr. John Stewart, Dr. Christoph Stoll

This is a revised version of the article first published in: F. Austermühl, J. Kornelius (eds.). *Learning Theories and Practice in Translation Studies*. Trier 2008.