
E-terminology*

Mariëtta Alberts, *Systems Development and Research, National Language Service, Department of Arts, Culture, Science and Technology, Pretoria, Republic of South Africa (vt05@dacst5.pwv.gov.za)*

Abstract: This article addresses computer applications in terminology and terminography. Terminologists aim at creating source language terms and target language term equivalents for new concepts and thereby enhancing communication in various subject areas and domains. The worldwide market for terminological products is gradually expanding. There is an ever increasing need for the elaboration of terminologies and the distribution of terminologically relevant information, either in computerized form or in printed form, to provide available terminological data to as many users as possible. This article provides various strategies for different users to access multilingual terminology resources. On-line availability of terminological data can empower South African language practitioners and subject specialists. Matters dealing with terminological management systems form a vital part of this article. Outcomes, advantages, constraints and requirements regarding e-terminology are given.

Keywords: COMPUTER APPLICATION, ELECTRONIC TERMINOLOGY, E-TERMINOLOGY, HUMAN-COMPUTER INTERFACE, INFORMATION, KNOWLEDGE, KNOWLEDGE-BASED SOCIETY, LANGUAGE BARRIER, MACHINE TRANSLATION, MULTILINGUAL TERM EQUIVALENTS, MULTILINGUAL TERMINOLOGY, TERMINOLOGY APPLICATION

Opsomming: E-terminologie. Hierdie artikel behandel rekenaartoeappings in terminologie en terminografie. Terminoloë stel hul ten doel om brontaalterme en doeltaalekwivalente vir nuwe begrippe te skep en daardeur kommunikasie in verskeie vakgebiede en terreine te verbeter. Die wêreldwye mark vir terminologiese produkte brei geleidelik uit. Daar is 'n steeds groter wordende behoefte aan die uitbouing van terminologieë en die verspreiding van terminologiese relevante inligting, in óf gerekenariseerde óf gedrukte vorm, ten einde die beskikbare terminologiese inligting aan soveel gebruikers moontlik te voorsien. Hierdie artikel bied verskeie strategieë aan verskillende gebruikers om toegang tot meertalige terminologiese bronne te verkry. Intydse beskikbaarheid van terminologiese inligting kan Suid-Afrikaanse taalpraktisyne en vakkundiges bemagtig. Aangeleenthede wat met terminologiese bestuurstelsels te make het, vorm 'n integrale deel van hierdie artikel. Uitkomstes, voordele, belemmerings en vereistes betreffende e-terminologie word aangebied.

Slutelwoorde: ELEKTRONIESE TERMINOLOGIE, E-TERMINOLOGIE, INLIGTING, KENNIS, KENNISGEBASEERDE SAMELEWING, MENS-REKENAARKOPPELVLAK, REKE-

* This article is a slightly adapted version of a paper presented at the Sixth International Conference of the African Association for Lexicography, organised by the Sepedi National Lexicography Unit, University of the North, Pietersburg, 2-4 July 2001.

NAARTOEPASSING, MASJIENVERTALING, TAALHINDERNIS, TERMINOLOGIETOEPASSING, VEELTALIGE TERMEKWIVALENTE, VEELTALIGE TERMINOLOGIE

1. Introduction

The terminology of each subject field or domain is increasing with each new invention. Information is distributed and knowledge is acquired by means of terminology. The TermNet slogan says it all: *No knowledge without terminology*. Terminology can therefore be regarded as one of the major language resources to assist with communication. The availability of multilingual and polythematic terminology can be an indicator for estimating a nation's scientific, technological and economical achievements and social, cultural and educational development. In the information- and knowledge-based society of the 21st century, knowledge and information are commodities that should be available to the community at large.

Terminology applications cover aspects such as knowledge and language engineering, machine translation, multilingual markets in education, business, science and technology. At present e-commerce (electronic commerce) is a matter of course. There is no reason why terminology practice cannot conduct most of its functions electronically. This article touches on various aspects regarding e-terminology (electronic terminology).

2. Terminological data collections

Terminology has its best tool in knowledge development through terminological and knowledge databases. Science and technology based on advanced knowledge continue to increase in both quantity and quality. In the lively scientific, technological, economic, political, social and academic environments, large numbers of documents are continuously produced. In order to ensure creative and efficient future research, stakeholders have to process a huge amount of information across a wide range of disciplines and domains. Most of these documents are written in languages foreign to the speakers of the indigenous languages in South Africa and many of these have to be translated into the various South African languages. Such an increase in foreign literature, coupled with the limited number of trained translators available, have led decision-makers to seek labour-saving, cost-effective solutions to the translation process. At the same time, developed and developing countries do business with South Africa and documents therefore also need to be translated into foreign languages.

In order to overcome the many difficulties involving language barriers, more attention should be given to the enormous task of the terminologist in supplying multilingual term equivalents for the various new concepts in different subject areas and domains. E-terminology can solve many communication

problems.

3. Computer applications in terminology

The constant generation of knowledge and the fast pace of technological development have created the need for understanding different knowledge organisations in each specific area through the use of technology. Specialised communication has a central hub in terminology.

The number of computing and communication systems is projected to grow continuously within the next few years. From a prospective point of view new applications of the National Termbank will increase. The combination of knowledge transfer, technology and specific knowledge with terminology is expected to create a national information structure situated in the National Termbank, that will enable new applications (e.g. translations, machine translations, etc.), and will open a new set of specialised communications. This will have an impact on the academic, business, and educational world, by advances making human-computer interfaces with environments available. Tools to support those with disabilities (e.g. the blind (Braille; speech recognition), the deaf (e.g. illustrations of the various signs of a sign language) or the handicapped (e.g. adapted keyboards)) will broaden the reach of the computer and communication revolution.

4. Standardised access to lexical and terminological data

In today's fast-paced global e-marketplace, a streamlined multilingual translation workflow will present government departments, institutions and companies with a solid competitive advantage. Standardised terminology contributes significantly to the quality of translations. As government departments, institutions and companies strive towards globalisation, consistent multilingual and polythematic terminology becomes a critical issue.

5. The National Termbank

The Terminology Coordination Section of the National Language Service, Department of Arts, Culture, Science and Technology, is the only language office in South Africa that devotes all its energy and time to terminology work. As the national office for terminology work, its aim is to coordinate all terminological activities in South Africa.

The Terminology Coordination Section renders a terminology service for the advancement of the official languages on a national level. Its vision is to develop, document, standardise and disseminate multilingual terminology to promote and facilitate communication in various subject fields and domains. For this purpose the Terminology Coordination Section maintains an electronic

National Termbank. Terminology work is done in close collaboration with stakeholders, subject specialists, linguists, academics, end-users and business partners (such as the national, provincial and local government, PANSALB and its structures, e.g. the National Lexicography Units (NLUs), the National Language Bodies (NLBs), Provincial Language Committees (PLCs), etc.).

In the 1950s, terminologists started documenting terminology on index cards. In the 1980s, the Coordinating Terminology Board (COTERM) initiated research on various aspects regarding the computerisation of terms and on the terminological management process. After thorough investigations, the national terminology office in 1996 decided to purchase licenses to use the MultiTerm database system. MultiTerm is a professional multilingual terminology management system developed by TRADOS. At present most of the functions and activities of the Terminology Co-ordination Section can be carried out via the electronic media.

In the Terminology Coordination Section, data is usually excerpted manually from documentation, but is immediately documented in the computerised MultiTerm database program. Terminological data that is obtained electronically from discs or compatible programs (such as the Translator's Workbench) can be entered into the database itself. Terminological data created by means of the TRADOS Translator's Workbench is automatically captured in the MultiTerm database.

The TRADOS Translator's Workbench for Windows represents state-of-the-art as far as powerful and flexible management of linguistic reference material is concerned. It offers access to previous translations not only at word level, but also at sentence level. It therefore helps translators who want to be relieved of time-consuming and repetitive tasks.

The Translation Memory (TM) of the Translator's Workbench is a database in which source language sentences are stored together with their target language equivalents. An additional system, WinAlign, matches source language terms or sentences with their target language equivalents. The memory learns unobtrusively while the translator works, and ensures that no repeated term or phrase in this database has to be kept twice or is used contradictorily. Fuzzy searches in the database give the user instant reference to text segments already translated. It can also manage additional information such as the client of the translation project or the domain in which translation units were created. The additional data can later be used to distinguish between different subsets of the data stored in the memory.

Active terminology recognition is one of the facilities built into the Translator's Workbench system. This system functions in collaboration with the MultiTerm database. Known terms already stored in the MultiTerm database are automatically highlighted on the word processor by the Translator's Workbench system. At the click of a mouse button or a keystroke, their translation equivalents can then be pasted into the document being translated. Compound words, morphologically modified forms, and even parts of sentences can be

found with a fuzzy-matching algorithm, even if they do not occur in the same form in the text to be translated. New source language terms and multilingual equivalents can be added to the terminological database through the Translator's Workbench.

The Translator's Workbench gives instant and flexible access to previously translated material because of its translation memory and the concordance features, and assists with active terminology recognition performed in cooperation with MultiTerm.

MultiTerm is integrated into the Translator's Workbench but is also an independent program. This means that terminologists can work on MultiTerm while a translator can also access the terminological database, even when he/she is not working in the Translator's Workbench.

By collecting multilingual and polythematic terminology, the Terminology Coordination Section builds very large corpora on specific subject areas and domains. From these corpora, multilingual terminology lists, technical dictionaries or CD-ROMs can be compiled for dissemination to end-users. Collaboration with related professional and academic institutions, subject specialists and linguists promotes quality control and the standardisation of terminologies. The Terminology Coordination Section aims to coordinate all terminological endeavours in South Africa and to be a clearinghouse for all terminological activities by sharing terminology and terminological knowledge with a multilingual content (see Diagram 1).

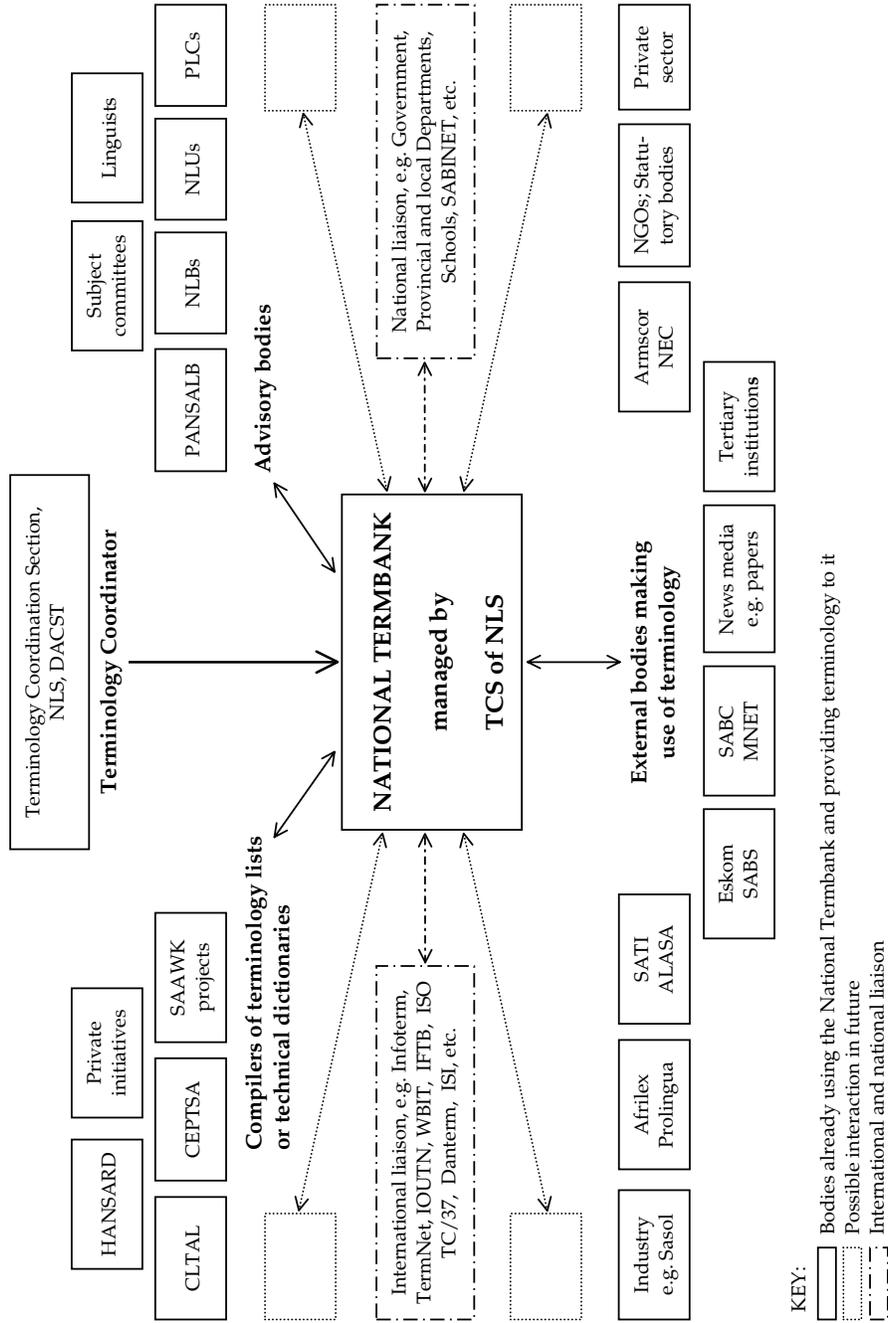
6. Terminology management on the Internet

Knowledge and information can only be transformed into a competitive differentiator when deployed and distributed wisely. Proper terminology management supplies such a tool. A terminology office that focuses on electronic devices as solutions for terminology management, capitalises on the added flexibility and efficiencies offered by proper client-server architectures and Web hosting in particular (Sanders 2001).

According to Crespel (2001), it is possible to provide a central and organised access point for the linguistic sector and to build and develop a linguistic portal to cover the complete range of multilingual activities, disciplines and needs of the sector. The set of objectives offered by a portal should be organised under the following sections:

- On-line consultation of linguistic resources:
 - all kinds of resources: terminology, lexicography, textual corpora, registration of terminology projects, etc.;
 - multilingual resources: no language limitation.
- On-line acquisition of linguistic resources, tools and services:
 - the possibility of viewing and/or testing the resources and tools before acquisition;

Diagram 1: Terminology Coordination



- needs assessment studies and results, e.g. priorities;
- on-line payment mechanisms.
- Technical, scientific and business information:
 - links to existing sites containing newspapers, publications, etc.;
 - "jobs on-line" directory;
 - market players' directory;
 - support for publications of research projects.
- Forum services:
 - inventory and links to existing newsgroups;
 - moderated newsgroups;
 - implementation of specific chats.
- Value-added services:
 - resource creation, validation and adaptation support;
 - consulting services;
 - project development analysis, publications of bids for tenders, etc.

7. Terminology dissemination

The focus of terminology dissemination is shifting from traditional dictionary publishing to efforts of content provision in termbanks. There are new challenges regarding terminology dissemination.

Implications for data management and the marketing and dissemination of terminology are the following:

- new challenges and related specialist dictionary publishing needs with regard to
 - terminological sources, and
 - the validation of dictionary entries;
- quality criteria when managing large heterogeneous terminographical contents;
- transition in the data capturing process;
- a single databank for multiple print and electronic devices;
- globalisation of the market;
- the necessity for co-operating with experts in the field such as information scientists, software developers, terminologists and publishers;
- business strategies to address the consumers' (dictionary users') present needs;
- new co-operation schemes for the dictionary publisher as content provider via intranet and Internet solutions with other partners;
- a considerable requirement for standardisation in the field of lexicography for the harmonisation of terminographical and lexicographical resources to enable data exchange (SQL configurations);
- a successful terminological data dissemination and an opening of new profitable markets without hidden threats concerning quality content

and protection of intellectual property.

8. Convergence of technologies and language engineering

Terminological practice is moving into a world where computers of all kinds, television, the Web, telephones, music media, graphic design and electronic books are evolving toward a seamless and yet heterogeneous environment in which information flows across platform boundaries to guarantee instantaneous terminological access anywhere anytime (Wright 2001).

Terminology is also in the process of becoming mobile: m-terminology is possible in an era of mobile computing (m-computing) and mobile commerce (m-commerce). The electronic means of creating, sharing, re-using and coordinating terminological knowledge for multilingual and transcultural content management paved the way for mobile terminology. It is already possible for terminology users abroad to access various termbanks by means of personal diaries and mobile phones (cell phones). Although this possibility has not been introduced in South Africa yet, there is no reason why it cannot be done. In South Africa one can already conduct business via cell phones (e-commerce) and one can listen to broadcasts from a radio station (e.g. Radiosondergrense) on the cell phone. It is therefore possible to connect the National Termbank with mobile devices to enable access to the extensive terminological variety.

9. Envisaged Web-enabled management model (TRADOS MuWA and TermCo)

The NLS aims to make the terminological information in the National Termbank available to all end-users via the Internet. The NLS started utilising the Trados MultiTerm system in 1996. With this version of the program terminological data can only be displayed in flat files on the Internet. An updated version of MultiTerm with an Internet interface has been available since the end of 2001 and the NLS system has been upgraded to enable users to consult the live data in the National Termbank.

9.1 TRADOS MultiTerm Web Access (MuWA)

The TRADOS MuWA form is used together with MultiTerm to present organisations with access to the National Termbank on the Internet or intranet for wider use. It fits perfectly with the current use of MultiTerm by the NLS.

MuWA offers a robust infrastructure for distributing, accessing and managing terminology over the Internet/intranet. Using a standard Web browser (such as Microsoft Internet Explorer or Netscape Navigator), users can be granted instantaneous, effortless access to the contents of the National Termbank. The termbase is hosted on an SQL database, which is accessed by the cli-

ent browser through the Web server.

MuWA enables centralised administration and management on the Web server. This administration model offers significant benefits to language workers by giving them access to the multilingual and polythematic terminology available in the National Termbank. It also presents mobile engineers or travelling executives with a valuable multilingual communication tool.

The read-only client access enables the terminology administrator to have complete, real-time control over content and revisions.

9.2 TermCo

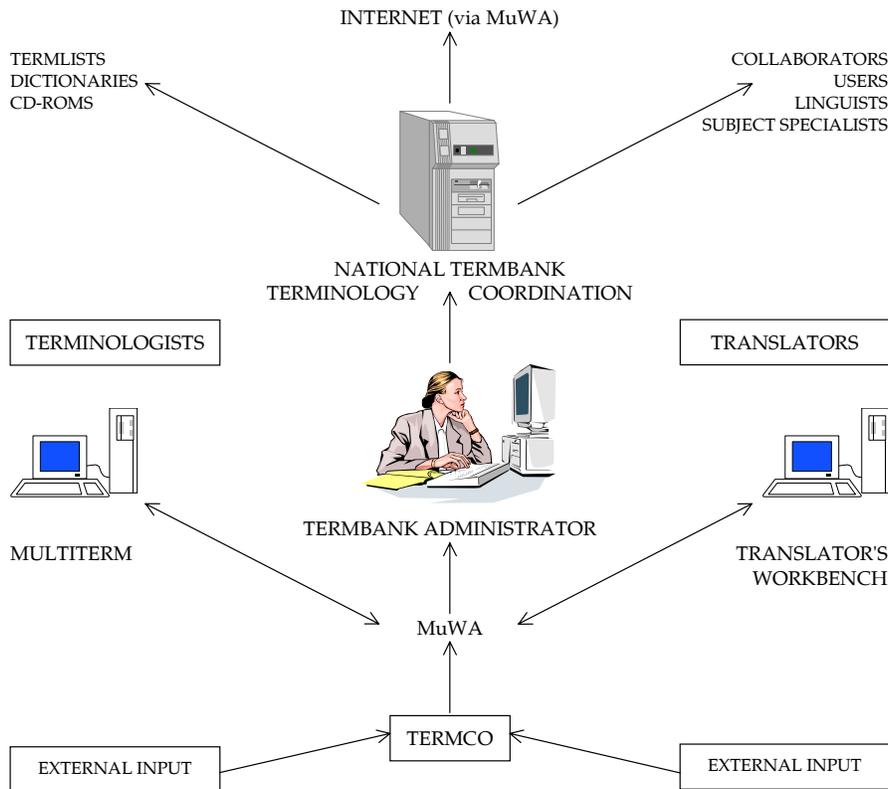
The Unit for Language Facilitation and Empowerment at the University of the Free State holds the TRADOS distribution license in South Africa. This unit recognized the need for terminology contributions to the National Termbank by terminology users. These contributions will have to proceed via an interface between MultiTerm and the Internet. The NLS commissioned the Unit to do a pilot study to establish the actual Internet-based software for a Web-based terminology management system. The idea with this pilot study was to showcase the system's functionality, and a detailed system design will follow to ensure the full and successful implementation of the system.

The envisaged model is based on the expansion of the existing electronic management of terminology by the NLS with TRADOS MultiTerm. The easy use and benefits of the Internet make this model the obvious choice to incorporate as part of a terminology management system. Although not all terminology users have access to the Internet, this is a valuable way of disseminating terminological information. Users may find access through normal information technology support channels, e.g. at libraries.

The envisaged Web-enabled Terminology management model consists of the following components which, when combined, form the model schematically represented in Diagram 2:

- The National Termbank is managed by the Terminology Coordination Section at the NLS.
- Queries are made via TRADOS MultiTerm Web Access (MuWA).
- The user accepts or rejects the terminological information.
- Contributions are made via TermCo or through other methods (e.g. e-mail, fax, telephone) in cases where the information is not available in the National Termbank or the user is not satisfied with the terminological information.
- Data manipulation is done by the terminologists.
- Data is available in the National Termbank.

Diagram 2: The Envisaged Web-enabled Terminology Management Model



9.3 Electronic terminology contributions to the National Termbank via the TermCo interface

The biggest disadvantage of MuWA as used in the context of the envisaged Web-enabled Terminology Management Model is the fact that MuWA only allows query facilities. With the TermCo Interface users will be able to comment on and supply new equivalents for consideration as additions to the National Termbank.

Persons (such as language workers, linguists and subject specialists) who will use and contribute to the envisaged Web-enabled Terminology Management Model will largely determine the success or failure of the model. The contributions will assist in the development of multilingual polythematic terminologies, which will enable the official languages to become functional languages

in all spheres of life. The Terminology Coordination Section of the NLS forms the hub around which the envisaged interface model will turn. The databank manager and the terminologists will actively manage the envisaged model and approve all contributions for inclusion in MultiTerm, MuWA and the National Termbank. There is a dire need for terminology co-ordination and the dissemination of terminological information in South Africa. This model enables proper co-ordination and sharing of existing terms and contributions. The model can also contribute towards the establishment of a multilingual society.

9.4 Manipulation of data

The Terminology Coordination Section of the NLS is the national coordinating body for terminology and will, through liaison with various stakeholders, ensure that the terminology entered in the National Termbank for dissemination is approved by the various governing bodies such as the different PANSALB structures, i.e. the 14 National Language Bodies (NLBs), the 11 National Lexicography Units (NLUs), the 9 Provincial Language Committees (PLCs), the national, provincial and local language offices, the Hansard offices, other linguists and subject specialists. The Terminology Coordination Section aims to be a clearinghouse for terminology documentation, systematisation and dissemination.

On reaching the Terminology Coordination Section via the Internet, the terminological data will have to be manipulated before it can be entered into the National Termbank. The terminology will have to go through the normal processing procedures and will have to be accessed by terminologists, subject specialists and linguists.

Various information fields are available in MultiTerm. When a contribution is made via the TermCo interface, the termbank manager will know who contributed and whether the contribution was made locally, from neighbouring countries or from abroad. Terminology for instance from Lesotho where different spelling and orthography rules for Sesotho apply, will have to be scrutinised by local Sesotho linguists and the National Language Body for Sesotho before it can be disseminated via the National Termbank. When Setswana terminology for instance is contributed, the terminologists responsible for Setswana will have to ensure that first language speakers of Setswana from the North West, the Free State and the Northern Cape and even from Botswana are consulted to ensure consensus on the term equivalents provided.

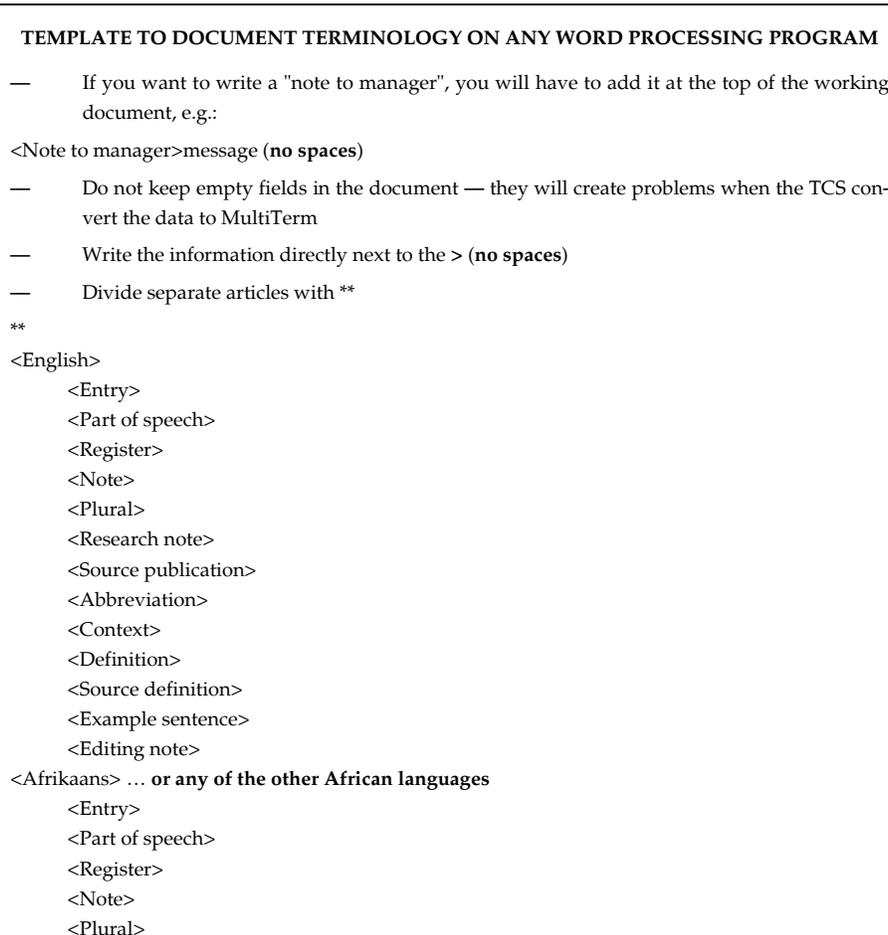
9.5 Contributions via word-processing programs

This interface cannot be implemented as yet. A new version of the Trados MultiTerm program has only been available since the end of 2001. The NLS will now be able to upgrade its terminological management system. In the mean-

time terminology contributions can be made electronically.

People who want to contribute terminology to the National Termbank and who do not have Internet access may also make their contributions electronically by using any standard word-processing software (e.g. Microsoft or Word-Perfect). They have only to use a template (see Diagram 3) supplied by the NLS with an indication of the various information fields to be presented. The terminological information can then be entered next to the relevant information field and a disc with the information can be presented to the Terminology Coordination Section for further manipulation of the data. The data can be utilised without being retyped.

Diagram 3: Template



<Research note>
<Source publication>
<Abbreviation>
<Context>
<Definition>
<Source definition>
<Example sentence>
<Editing note>

**

Contributions can further be made via purpose-specific lexicographical or terminographical databases such as Omnipage which is being used by the National Lexicography Units. Since MuWA and Omnipage are both SQL-based, there would be no problem to exchange data. Any other lexicographical or terminological database program with a SQL configuration will enable data exchange.

10. Outcomes regarding e-terminology

- Linguists, language workers, subject specialists, laypeople, students and learners have easy access to multilingual, polythematic terminology.
- Terminology in original and translated documents can be documented by utilising the TRADOS Translators' Workbench system. Terminology can be excerpted from source language documents and the translated versions in the various target languages. With the aid of the WinAlign program the source language term and target language term can be aligned for documentation in the MultiTerm database.
- Terminological corpora can be used for the parsing of language systems and the development and testing of machine translation programs.
- Multilingual terminology in the National Termbank can be used for machine translation (e.g. the Lexica program).
- Existing published dictionaries, which may already be out of print, can be recirculated by means of CD-ROM publications.
- Terminology can be used for artificial intelligence purposes (e.g. speech recognition).
- The various concepts represented in the National Termbank can be utilised by the deaf or hearing-impaired community since diagrams or illustrations of signs or pictures of the signs can be included in the MultiTerm database.
- Corpora can be built for utilisation by lexicographers (e.g. NLU's). The building of corpora is a very important basic data-creation instrument in an information society. It is considered as the first level infrastructure of language information. A parallel corpus that shows the co-relations between two or more languages can be used for machine translation,

-
- multilingual information retrieval, foreign and indigenous language training and comparative philology.
 - Spelling and orthography rules can be tested against the available terminological data (e.g. by NLBs).
 - Spelling control systems can be developed by utilising the electronic data.
 - Studies can be conducted to determine the word-forming and term-creation principles of the various languages.
 - On-line terminological data will be available via mobile devices (e.g. cell phones and personalised diaries).
 - The advantages of on-line access via mobile hand-held devices:
 - they are designed for personalised information;
 - they almost always accompany the users;
 - they are almost always switched on;
 - they make terminological information accessible anytime;
 - users have access to them in real time (updated information is available).
 - Constraints:
 - their screen-size, input capabilities and bandwidth are limited.
 - Requirements:
 - interaction would have to be minimalised;
 - short, meaningful headlines would have to carry relevant information;
 - the screen lay-out would have to be simplistic.
 - Electronic language management through human language technology (HLT) can benefit from e-terminology (e.g. speech recognition, signal recognition, multilingual speech processing, automatic language identification, machine translation, automatic analysis, multilingual information retrieval and computer assistance in text creation and editing).

11. Conclusion

Given the rapid development of economy, culture, science and technology in the world and the value of terminology as communication resource, it is becoming increasingly important to document terminology in such a way that terminological information can be easily retrieved. The multilingual and polythematic terminology should be collected and documented systematically and managed in a coordinated way for proper distribution, publication and application in language and knowledge engineering. The input of collaborators can empower the Terminology Coordination Section to standardise available terminology and to give access to high-quality and reliable terminology. The advantages of e-terminology can promote the terminology management process, the infrastructure and the communication process.

Bibliography

- Baca, M.** 1997. *International Terminology Working Group Minutes, September 5, 1997*. Paris: Offices of the Inventaire Général.
- Crespel, Alexis.** 2001. *GEMA: Gates for Enhanced Multilingual Resource Access*. Paper read at the Fifth TermNet Symposium TAMA 2001, 1–2 February 2001, Antwerp, Belgium.
- Derouin, Marie-Jeanne.** 2001. *From Traditional Dictionary Publisher to Content Provider. Implications for Data-management and Marketing in a New Publishing Era*. Paper read at the Fifth TermNet Symposium TAMA 2001, 1–2 February 2001, Antwerp, Belgium.
- Nyangintsimbi, L.L.T. and J.H. Hattingh.** 2001. *Report on the Findings of the Terminology Survey at the Provincial Language Units (TermCo Project)*. Bloemfontein: University of the Free State.
- Plested Avarez, María Cecilia, Díaz Castrillón and Adriana Lucía.** 2001. *Terminology in the Information Age — Sharing Terminological Knowledge. Colterm Databank: From Concept Systems to Terminological Entry Management*. Paper read at the Fifth TermNet Symposium TAMA 2001, 1–2 February 2001, Antwerp, Belgium.
- Sanders, G.** 2001. *Sharing Terminology: The Benefits of a Web-based Workflow*. Paper read at the Fifth TermNet Symposium TAMA 2001, 1–2 February 2001, Antwerp, Belgium.
- Sonneveld, Helmi.** 1997. *Terminological Principles and Methods in the Translation of Thesauri. International Terminology Working Group Minutes, September 5, 1997*. Paris: Offices of the Inventaire Général.
- Wright, Sue-Ellen.** 2001. *Convergence Technologies and Language Engineering*. Paper read at the Fifth TermNet Symposium TAMA 2001, 1–2 February 2001, Antwerp, Belgium.