Interaction of Computer and Language through Computational Terminology

Georgiana Tofan
Ștefan cel Mare University
Suceava, Romania
georgiana.tofan@usv.ro

Abstract: To understand relationship between computer and language, this article attempts an overview of computational terminology, including various linguistics aspects generated through electronic medium. The advent of each new computational stage provides different language situations, whose functionality in new roles implies linguistic analysis, such as for terminology extractions or electronic discourses. Concomitant with the approach to aspects of computational terminology, we intended to emphasize the importance of choosing the appropriate technical and linguistic methods, in order to acquire efficiency in established contacts between computer and language. As a result to Internet communication, Netspeak comes to capture the aspects of electronic world, according to interactions between users.

Keywords: universal language, Internet culture, computational terminology, e-discourse, Netspeak.

1. The language of the Internet

The emergence of different communication technology heralds evolutionary changes which bring feelings of excitement and trepidation. Although the computer is "a machine capable of executing instructions on data"¹, or "an electronic device for storing and processing data"² as the dictionaries define that word, still this is not the reason why we are part of an "Information Era". That reveals other important aspects which connect computers to our lives: storing data, managing complex communication environments, processing images and sounds, and so on. Computers have proven to be very useful tools in academic, industrial, government environments, their efficiency being recognized in almost every activity field.

The increasing use of the Internet, considered to be the world's largest computer network, is having general and individual effects on language, as a direct result of the new ways of communication, substantially different from other semiotic acts.

Before the Internet adventure begins, there were only two ways to send and receive information: by reading or verbal communication. The nature of the electronic medium, along with the Internet technologies, contributed to the dissemination of information among people all around the world, becoming widely available.

It is certainly the case that the situations of Internet communications involve linguistic consequences, once the individuals who take part of this do not share the same culture. A similar point is made by David Crystal in Language and the Internet, where, starting from the notion of "global village" assigned to the Internet community, he says: "There has been much talk of the notion of a "global village", which is at first sight a persuasive metaphor. Yet such a concept raises all kind of linguistic questions. A village is a close-knit community, traditionally identified by a local dialect or language which distinguished its members from those elsewhere (...) If there is to be a genuine global village, then we need to ask 'What is its dialect?', 'What are the shared features of language which give the world community of users their sense of identity?' And, if we cannot discern any unifying dialect or language, or a trend toward such a unity, we need to ask

---

ourselves if this “global village” is anything more than a media fiction. To explore the way in which the concept of “global village” finds its linguistics expression, we need to take into account the global spread of English, that seems to affect the existence of million of people. This is meant to highlight the fact that, nowadays, English is the world’s leading language, being used in many countries: “It is used on all continents. In surprisingly many countries (more than 100, according to recent estimates) it has important internal functions as “Second Language” in addition to one or more indigenous tongues, being used in politics, business, education, technology, the media, etc”. This is an indicator that might establish the existence of a universal language, whose functions, at the level of the Internet medium, improve conditions for global communication, and turns out to be highly successful.

All these observations need to be related to the growth of the Internet culture, a worldwide phenomenon that reflects different backgrounds of the users, constrained by their status to confront various Internet situations.

2. Computational Terminology: Needs and Purposes

As a discipline, terminology is concerned with the study of the terms, their formation, usage, significance, evolution, linguistic aspects of the terms representing a major interest to terminologists and specialists, as well as to the translators. In addition, researchers from different fields of study have strong interests to develop and maintain a terminological rigor, in order to provide a precise understanding.

In terms of needs, the task of creating terminologies is clearly illustrated in Recent Advances in Computational Terminology, edited by Didier Bourigault, Christian Jacquier and Marie Claude L’Homme: “Information science specialists, lexicographers, translators, as well as other specialists, such as engineers, scientists, and technicians have experienced the need to improve communication, or to access subject specific information. Several types of terminological (or terminology-related) resources, such as specialized dictionaries, term banks, glossaries, and thesauri, have been built in order to satisfy these needs”. Following the above mentioned authors, we understand that such needs would be an important source of purposes in computational terminology, a field that requires both theoretical and practical approaches, whose implications lead to “identification of terminological units” (referring to the automatic or semiautomatic ways of linguistic approaches at the text level) and finding the “semantic and conceptual information on terms” (the linguistic operation after the terms have been extracted).

2.1. Aspects of automatic term extraction

The automatic extraction of terms from specialized texts was a real challenge for terminology extraction (TE), motivated by the large number of textual corpora in language processing, as from the moment of appearance of TERMINO (term detector) in 1990, specialist showed their interests in development of software tools specialized in terminological field.

Although there are different approaches for automatic term detection and extraction, the demands of creating this type of programmes require common type of information, as they create possibilities of dealing with electronic texts.

Systems of terminology extraction exemplify the potential of computer technology, using linguistics elements; in fact, the field is wide open for new initiative, inspired by such mechanisms. Based on this view, M. Teresa Cabrè Castellvi, Rosa Estopà Baquot, Jordi Vivaldi Palatresi describe some systems of terminology extraction, pointing out their positive and negative features. We will discuss the architecture of the semiautomatic extraction system, called ANA (Automatic Natural Acquisition), having as components two modules: “a familiarity module” and “a discovery one”. The first module is based on three groups of words, while the discovery module,
representing the process of creating the semantic acquisition, (whose main goal is to detect and generate terms) depends on three types of interpretation of the identified words. The construction of the familiarity module contains the following groups of words:

a. function words (i.e. empty words): a, any, for, in, is, of, to...

b. scheme words (i.e. words establishing semantic relationships) such as box of nails, where the preposition shows some kind of relationship between box and nails.

c. bootstrap (i.e. set of terms that constitutes the kernel of system and the starting point for term detection)⁹.

In other words, this familiarity module has an important role in the extraction system, and can be expressed as the predefined part of it. Furthermore, the discovery module involves generated terms, serving as an identifier to other activity (first module); the second module can be seen as a term selection linked to three types of word interpretation: expression, candidate, and expansion. Term expression ($T_\text{Exp}$) deals with the frequency of the existing terms, every new word being added to the semantic network; term candidate ($T_\text{Can}$) represents the situation when an existing term is framed to be a new word inside the system, because it appears next to another word; term expansion ($T_\text{Exp}$) assesses terms identified in the same word sequence. But, for providing an explicit view, we will illustrate the previous types of word interpretation of the involved system, making use of the following examples and a figure:

- **expression** – For instance if the system has diesel and engine as known terms and find sequences like: ... the diesel engine is...or...this diesel engine has...Then the sequence diesel engine is accepted as a new term and is included in the semantic network as a new node with links to diesel and engine.

- **candidate** – an existing term appears frequently together with another word and a scheme word as in:... any shade of wood...or...this shade of wood... or this shade of colour... Here shades becomes a new term and is placed in a new node of the semantic network.

- **expansion** – an existing term appears frequently in the same word sequence, without including any scheme word: ... use any soft woods to... or... this soft woods or... As a result, soft wood is incorporated into the term list and the semantic network as a new node with a link to woods.

![Figure 1. Term candidates interpretation.](image)

### 2.2. The evaluation of the ANA

The functionality of the system presented requires a minimum of linguistic resources, aspect that facilitates an accessible usage, but, this is not the only condition for a positive feedback. In fact, however, this system does not assure a completed validation of the terms, and, more importantly, it gives the possibility for non-valid terms to be included to the term list. Still, it is quite natural to think that, despite all negative features, this system could be a pretext for improving computer technology, to create new mechanisms of (semi)automatic term extractions.

Perhaps ANA greatest importance is its existence, as to improve terminology extraction, a starting point is needed. Broadly speaking, the efficiency of this type of systems was not clearly measurable. Under these circumstances, the progress in the field of computational terminology needs to combine in a more active manner technical and linguistic methods.

### 3. The linguistics aspects of the Internet situations

As more and more individuals become part of the Internet communion, almost everyone is familiar with e-mail, chatgroups or blogging, forms of electronic interactions, whose linguistic features depends on style, purpose, and audience.

Thus, the relations between computer and linguistics must be perceived as an important interaction, focusing on progress in developing softwares for linguistics: "Of course, one reason the computing metaphor works so well is that language truly is a form of software. Just as the human brain was the model for computer hardware, human language was the model for..."¹⁰

---

⁹ Iden, p. 57.

¹⁰ Iden, p. 58.
computer software — and we are now, after a
decade of widespread, intensive experience with
computers, in a position to recognize what that
means. 11

Linguistics and computer in various fields of
activity, including hardware and software
industries, promote their common goals, allowing
us to gather efficiently all kind of information,
the relation between linguistics and computer
involving a close cooperation between the two
types of users.

The language of Internet displays the process
of writing in its different forms, depending on
users' possibilities of expression, knowing that
the communicative problems within e-discourses
of all kinds could affect language using. But
there are no specific ways of predicting
language-using situations, as it is evident that
Internet users are still in learning situations. As a
result, a conventional Internet behaviour, in the
sense of general agreed models, is required to
avoid linguistic problems.

The larger the number of Internet users is,
the various the situations become. Here,
therefore, we find stylistic approaches that
identified five main features for written language
(also found in Internet written situations):
graphic, orthographic, grammatical, lexical, and
discourse features, and another two features for
the spoken language (identified on the
Internet through songs, videos, films): phonetic
and phonological features.

> Graphic features appear inside written
language as different typography, font, size,
alignment, colour, contrast, white space, along
with other elements which define the creation of
the pages.

> Orthographic features (also called
graphological features) are the specific aspects
of using the alphabet, according to every
language, many ways of expressing being
considered (spelling, punctuation, etc.).

> Grammatical features expose various
systems of languages, and thus, elements of
morphology (word structure) and syntax
(sentence structure).

> Lexical features set clear distinction
between vocabularies of different languages or
even varies within one language.12

> Discourse features are based on
cohesion (organization of the discourse), and
coherence in terms of logical exposition of ideas,
relevance.

Spoken language on the Internet has a
strong presence through videos, songs,
webcams, because of the continuous
technological development. We are encountering
spoken language situations, and phonetic and
phonological features can be easily observed as
follows: phonetic features are connected to the
particular series of tones in the human voice,
while the phonological features deal with sound
systems of language.

4. Netspeak: a general view

"Netspeak" is a term coined by David Crystal,
made to represent Internet communication,
indicating that "speak" includes action of writing
as well; Internet language has to be represented
by specific features of the electronic medium, so
"Netspeak" comes to assume linguistic Internet
situations. It should be noted that features of
Netspeak were identified outside of the Internet,
influencing vocabulary in everyday
communication. As a matter of fact, computer
terminology comprises terms that were adapted
to the speaker's needs, so the special vocabulary
used by programmers was extended to
everyone's life. Accordingly, we hear all the time
"dot com (.com)" forward, slash, www", and
examples like "Let's talk offline (meaning let's
talk in private)" or "He is multitasking (in the
sense of doing two things at once)".

In the case of electronic interactions like e-
mail, chatgroups, Netspeak makes use of
expressiveness to send and even exaggerate
nuances of feelings, moods, ideas, through
symbols called smileys or emoticons, which are
created through combination of keyboard
characters. Beyond their expressive qualities,
smiley faces seem to complete language in an
innovative way, since written language has
always been in impossibility of using facial
expression. Sanderson involves even personality
features when he describes smileys, making the
following statement: "You might include a smiley
as a reminder of the ongoing context of the
conversation, to indicate that your words don't
stand on their own. A smiley can point out to the
other participants of the conversation that they
need to understand you and your personality in

---

11 John Lawler, Helen Arlott Dry, Using Computers in
12 David Crystal, Prosodic systems and intonation in English,
Cambridge University Press, Cambridge, 1969, in David
Crystal, Language and the Internet, Second Edition,
13 The presented situation refers to the concept of "language
variety", which represents language in different situations.

Terms like "register", "genre", are also used to reflect the
same idea.
order to understand what you've said\textsuperscript{14}. With an impressive collection of over 650 smileys compiled over the years, David Sanderson finds interesting meanings for computer characters (see the Table 1 below describing examples of smileys).

**Table 1. Examples of smileys (after Sanderson)\textsuperscript{15}**

<table>
<thead>
<tr>
<th>Basic smileys</th>
<th>Joke smileys</th>
</tr>
</thead>
<tbody>
<tr>
<td>:-) pleasure, humour, etc.</td>
<td>(-:) user is wearing a walkman</td>
</tr>
<tr>
<td>:{ sadness, dissatisfaction, etc.</td>
<td>8-) user is wearing sunglasses</td>
</tr>
<tr>
<td>;( winking (in any of its meaning)</td>
<td>B:) user is wearing sunglasses on head</td>
</tr>
<tr>
<td>;( ;( crying</td>
<td>:{ user has a moustache</td>
</tr>
<tr>
<td>%-( %-) confused</td>
<td>:) user is drunk</td>
</tr>
<tr>
<td>:o 8-o shoked, amazed</td>
<td>:l user is a vampire</td>
</tr>
<tr>
<td>:[ ] sarcastic</td>
<td>:E user is a bucktoothed vampire</td>
</tr>
<tr>
<td>:F user is a bucktoothed vampire with one tooth missing</td>
<td>7-: user has a cold</td>
</tr>
<tr>
<td>:@ user is screaming</td>
<td>:-) user is drunk</td>
</tr>
<tr>
<td>:-) user is a punk</td>
<td>;(- real punk don't smile</td>
</tr>
<tr>
<td>+:-) user holds a Christian religious office</td>
<td>0 :-) user is an angel at heart</td>
</tr>
</tbody>
</table>

| Smiley stories |  
| :o 8-) 8-) A smiley to disguise himself gets glasses and a fake moustache. |
| C:) >[ ] 8-) A smart smiley left watching too much TV. |

All these interactive ways of communication attempt to display users’ potential, facilitating a kind of language revitalization. As we have seen, the enormous expansion of the Internet determined a great creativity, enriching the community's linguistic norms.

Immediate innovation in the field of computational terminology has a clear impact on the kind of language we use. Many new technologies are anticipated, and arena of communication will always involve constant pressure for the adaptation to the new contexts. Internet observers therefore predict conversational openings, when we will have to get used to new models of communication.

Certainly, the participants of Netspeak seem to be aware of the fact that they are still in learning situations as long as communicative potential determines new rules and roles. The linguistic dimension will take time to evolve, in a field where technology continuous changes play a central role, especially among users forced to learn if they want to join this Internet community. When the survival and success on the Netspeak cannot be assured by the the simple use of language, the user becomes interested in the everyday realities.

The experience of language, along Netspeak is far from ending, so the computer-mediated language remains the core of interactions inside this environment.

5. Final remarks

Every computer contains information in some language, and data in language can be represented in different forms, according to application software. This is the reason why, as noted above, language and computer contact frequently support each other under the influence of advances in technological field. The fate of computer terminology may be determined by the evaluation through knowledge, in a phase of rapid evolution. It has been observed that the representation of computational terminology reflects important purposes for linguists implied in this field, in the construction of which language activities and results no doubt contribute.

Given this reality, then, a convenient starting point is represented by systems of terminology extraction. As long as a basis for these systems has been established, specialists have the advantage of building new mechanisms for (computational) terminlology. It might be pointed out that besides the semiautomatic systems for term extractions, Internet situations provide an appropriate environment for performing searches to obtain specialized terms. Text acquisition for terminological exploitation is a complex process that fell within the linguistics and computers, based on a conceptual structure. In computer medium, the question of language representation is accessed and applied through professional users, and widespread even at the level of common language.

Network utilization of terminology is an innovative method, that has helped identify and create new terms, and such projects must be undertaken to accomplish the demands of producing a terminology. In a phase of rapid evolution, computer terminology reports advances in both language and computer,
sustained by continuous emergence of new resources.

References


Sanderson, David, Smiley's, O'Reilly and Associates, Sebastopol, 1993.


Georgiana Tofan
PhD student, Ștefan cel Mare University of Suceava,
Faculty of Letters and Communication Sciences, PhD supervisor: Prof. Ioan Oprca, PhD.