The Need for Establishing a Terminology Database in Albanian as a Step towards Categorization of Terms

Linguistics

Keywords: terms, terminology, term databank, standardization, translation.

Dr. Arjan Shumeli
Dr. Luljeta Mine
MA. Esmeralda Sherko

English Department
Agricultural University of Tirana, Albania

Abstract

Establishing a terminology databank for both broad and narrow domains of knowledge is essential for the process of categorization and standardization of terms across different text sources and for translation purposes. The creation of this terminology databank is not a process that starts and ends simultaneously. On the contrary, the terminologist has to allow for a number of years between the time the new term is coined and the time the term is defined and elaborated before it enters this term bank. In addition, the terminologist has to examine closely the life cycle of the term to look for additional meanings being added to the original meaning or old terms dropping out of usage and being replaced by new ones. Neologisms are on the rise due to the unstoppable advances in technology and production and globalization is a reality amongst us. It is advisable to utilize a vast corpus of texts to create this database by using a concept-based approach to terminology as well as the appropriate computer tools and methods.

1. Introduction

Harmonization of terms, consolidation of terminology and establishment of equivalents amongst languages in various domains of knowledge is quite important. The establishment of terminology databases in the Albanian language is a task that has to be dealt with in a timely fashion, since Albania is a potentially candidate country and with its entry to the EU community arises the need for the establishment of a smooth communication between specialists and non-specialists in various domains of knowledge. The establishment of terminology databank in all fields including narrow domains of science and technology will ensure a sufficient supply of terminology in the Albanian language for translation requirements arising from the country gaining official status in the EU. In the European context efforts have already been launched by different countries including the new members (*Bulgaria, Romania*). An example worthy of mention is the *Inter-Active Terminology for Europe* which is a multilingual terminology database containing c. 1.5 million entries. There are c. 8.7 million terms in over 100 languages in the database, but the majority of terms (c. 8.6 million) are in the 23 official EU languages[1].

The work ahead for the terminologists and the specialists in the field is enormous. The established languages in the European institutions have to update and consolidate the entry of new terms in the IATE terminology, whereas the new languages of new entries into EU, including Albanian, as a candidate country, have to populate the IATE database with terms from various fields. In the future with all countries entering EU *specialist communication* is vital to sustaining both growth and the European fabric of society. The cooperation between countries in the EU will increase, people will be free to move around, multilingual communities will be created and the linguistic diversity will be enormous. The terms will be created and knowledge and information among people in different fields will be disseminated easily without linguistic barriers. The multi-lingual Europe will depend crucially upon the accurate translation of a wide range of documents, including scientific and technical documents.

[1] U. Bhreathnach., F. Cloke, C. N. Pháidín, "Terminology for the European Union: The Irish Experience: The GA IATE Project" Dublin, 2013.

2. What is a term and where to track a term?

The analysis of the composition of terminological lexicon of any given domain of knowledge and science from the *form* and *contents* perspective enables us to establish the contact and departure points between the *common word* and the *term*. A term, like the common word, serves to convey the expression of the many scientific and technical concepts[2]. It is easy to track down the meaning of those terms which have derived from the common (general-purpose) lexicon, since most of these words are reflected in a non-terminological dictionary. According to J. Thomai the word is connected to the meaning as a semantic-linguistic element, while the term is connected with a specialized meaning, which is stripped of it emotional and stylistic charge[3]. From this perspective, the term is used to convey a certain concept within a system of concepts[4]. Furthermore, the term retains the meaning within any context, that is a meaning that the term possesses in its system of concepts. This characteristics of the term helps us draw the dividing line between the term and the common word.

Each domain, be it *agriculture*, *mechanics*, *forestry*, etc., is differentiated by the linguists through the use of specific and specialized words and terms, which are used to disseminate knowledge among the community of specialists and experts. The community of agro-mechanists embraces experts in the fields, students studying agro-mechanics at the Faculty of Agriculture at AUT, terminologists and translators of different sources of materials in the field. It would suffice to analyze any given text or corpora of spoken discourse in the field of agro-mechanics to track down the domain and the specific terms and terminology used under that domain. The above categories use the so-called idiosyncratic words, *that is the terms*, but also words with certain restricted semantics. It is important to stress that such specialized terms reduce ambiguity among speakers of this community, because its members own and possess a specialist knowledge in the field. The number of terms which constitutes the agro-mechanics special language of this community is approximately 1900 terms. The attributes of the terms and their definitions vary in accordance with the type of text. In many cases the *headword* , as in '**plough**" is accompanied by a definition, its grammatical category, and some other references or characteristics: 1-Plough: *The plough (BrE) or plow (AmE; see spelling differences; /ˈplaʊ/) is a tool (or machine) used in farming for initial cultivation of soil in preparation for sowing seed or planting to loosen or turn the soil*[5]. But the typical term bank of "**implements**" may have more attributes showing a variety of relationships. The following graph is based on the typical contents of the term bank "*farm implements*" in Albanian.

[2] A. Duro, '*Terminology as a system*", Panteon, 2001, Tirana.
[3] J. Thomai," Change of semantic meaning in Albanian language", 1989, Tirana
[4] See Duro above.
[5] http://en.wikipedia.org/wiki/Plough

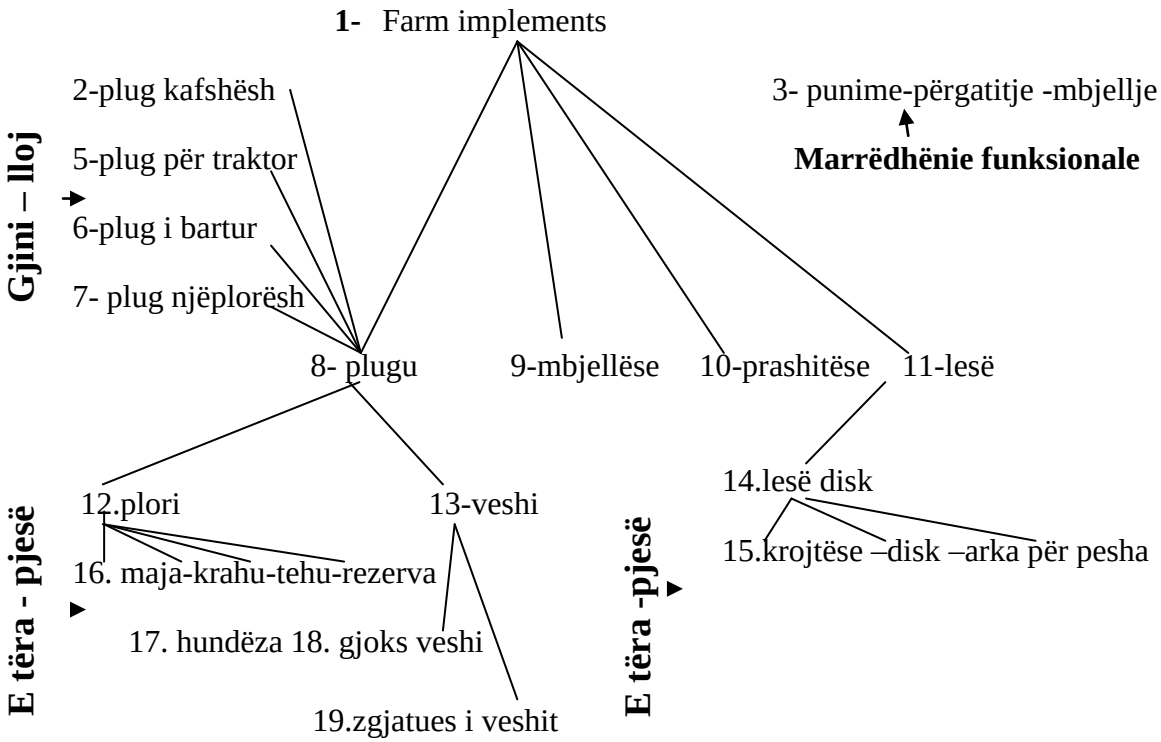

Fig 1. Representation of the relationships in an Albanian term bank

The number of terms which usually constitutes the special purpose vocabulary of a domain varies from 2,000 to 2,000,000. The attributes of a given term vary again: in a typical term bank there may be as few as 40 attributes or as many as 76. These statistics are based on the typical contents of a specialist dictionary or the number of terms in a term bank.

The term '**farm implement**" (alb. *vegel bujqesore*) is used on a frequent basis and as an exclusive definition in the field of agro-mechanics. The categorization of the term is done by depicting certain fundamental features that characterize it in various processes. The **farm implement** is used in the following usages: (1) soil cultivation (alb. *kultivim toke*) as **plow** (alb. *plug*); **cultivator** (alb. *cultivator*); **harrow** (alb. *lesë*); (2) planting (alb. *mbjellje*) as in **seeder** (alb. *makinë mbjellëse*); (3) fertilizing and pest controls (alb. *plehërim dhe kontroll demtuesish*); **manure spreader** (alb. *makinë për shpërndarjen e plehut*); **irrigation** (ujitje); **drip irrigation** (alb. *vaditje me pikim*) etc. All of the above usages or *nodes* make up the *hierarchy of the term* '**implement**". It should be noted that in each of the nodes we may end up with a cluster of associated or related terms specialized to carry our related processes as in **planting machine** (alb. *makinë mbjellëse*); **trailed planter** (alb. *makinë mbjellëse e tërhequr*); **seeder** (alb. *makinë mbjellëse fare*); **row-crop drill** (alb. *makinë mbjellëse me rreshta*); **distribution planter** (alb. *makinë mbjellëse me shperndarje*). In such word combinations it is important to look at the description of the relationships between parts and the whole as in the case of **plough** (alb. *plug*)", which is made up of the following parts: **mouldboard** (alb. *vesh plugu*); **share** (alb. *plor*); **landslide** (alb. *korpus*); tailpiece (alb. *zgjatues i veshit*).

3.What elements of the terms should a terminology database consider?

The establishment of the *terminology databank* should consider the entire *life cycle* of the *term*, hence the involvement of the terminologist and the domain specialist. The work of terminologist is of paramount importance. The terminologist has to examine closely the term, because, like any other living organism, the *term* is coined (*birth of the term*); it evolves over time (*evolution of the term*) by changing its semantic meaning to the point where the modern and contemporary meanings may be added to the original ones. Evolution of terms might also involve *widening of*

meaning, *narrowing of meaning*, *grammaticalization*, *terminologization* (a word rises to the level of the term), de-*terminologization*[6] (a term gaining the status of a common word)[7] etc. Each and every term may enjoy wide usage across various domains given a shared feature or characteristics, which can be added, removed, or altered over time, often to the extent that cognate terms across space and time have very close or different meanings. The term may also die and in such cases we may refer to the '*demise of the term*". The terminologist should be more focused on the current and contemporary definitions of the term. His narrow focus of work should be the *description* and the *maturity status* of the term. In considering *candidate terms* he has to look closely at the tools and methods by which the term is acquired. On the other hand, the domain specialists have to bring in different uses and instances of the terms from their own expert texts. They have to report new cases of term coinages expressing their criticisms and affirmations on the use of the terms. Their role is essential in defining and elaborating the newly-coined and established terms.

In analyzing the texts the terminologist and the specialist have to keep an open-ended class of terms. Unlike other domains of knowledge, like mathematics, chemistry etc., where the work of the terminologist is easier since these sciences have borrowed heavily from Latin and Greek, agro-mechanics, as a narrow domain, presents quite a few challenges. The terms in the field of agro-mechanics are as old as the Albanian language (*this relates to the spirit of Albanian being tied up with the framing of the land*), it has a native flavor and the terms derive primarily from the authentic basis of Albanian. Hence the work to trace the origin of these terms and how they evolved over time is quite huge and time-consuming. A close reference should be made to "**cart**" (alb. *qerre*) and "**plough**" (alb. *plug*), where the terms indicating the constituent parts of these ancient implements are quite innumerable. The work of terminologist is even harder because the community of specialists and non-specialists in the domain of agro-mechanics used to be pretty vast, given the status of the country as an agrarian–oriented country (1945-1990), where the vast majority of population dealt extensively in agriculture. This triggered the widespread use of agriculture and consequently the use of terms and word combinations among the common population. Such a process involves the *terminologization* and the *determinologization* of terms and their abundance. This has to do with the *penetration of the terms into the common lexicon from the so-called micro-languages of science and techniques*.[8]

Further to the above, the terminologist should sift through a wealth of expert and specialist documents, that is primarily materials which are written by experts and specialists in the field. The term in any given expert or specialist text is represented through an explicit meaning attached to it as well as with its corresponding grammatical characteristics. The selected term should record all of the attributes of the term. Another important step would be to depict the contextual characteristics of the term through the illustrative usage of the term in order to avoid ambiguity. The meaning of the term '**planter**" (alb. *makine mbjellese*) would be better served by providing the definition of the term "*Like a grain drill is an agricultural farm implement towed behind a tractor, used for sowing crops through a field*"[9]", as well as by providing the illustrative usage through an image. This would be helpful when you try to indicate the different types of a planter.

The terminologist should also watch out for *colloquial synonyms* which in the case of Albanian turn out to be abundant both from the linguistics and historical perspective. In such an endeavor efforts should be made to consider foreign language equivalents, since several languages enjoying a high status, like English, have already a standardized language. In short, all of the following elements need to be looked at when building up the terminology databank: acquisition of the term, representation of the term, and maturity of the term.

4. The need for an electronic computer-based terminology database in the field of agro-mechanics

Over the last 10 years computational linguistics has made huge advances. Various features and characteristics of terms can be processed through the computer system. The processing of language data has been more successful in the case of terminological lexicon. According to A. Duro[10] the use of computers is more productive with terminological lexicon, because this type of lexicon appears more consolidated and *workable*, hence it is much easier to be processed through different computer models. The use of computers allows for analyzing a vast wealth of lexical

[6] Th. Feka, *"Observations on terminological lexicon in the Albanian language dictionary*", Studies on Lexicon, 1983, Tirana
[7] A. Shumeli, "*The formation and functioning of the agro-mechanics terminology in Albanian as opposed to English*", PHD thesis, Tirana, 2013.
[8] A. Kostallari, Gjuha letrare kombetare shqipe dhe epoka jone", Tirane, SF, 1984.
[9] Merriam Webster dictionary Online
[10] A. Duro, Terminology as a system", Panteon, 2001. Tirana.

items and processing them through computer programs[11]. Many linguists, including Duro, use the term *language engineering*. To all outward appearances language engineering is the same as other types of engineering including mechanical and civil engineering. The two of them deal with the planning, design, production and generation of products. Whereas language engineering deals with the processing of building blocks of language. In addition its ultimate purpose is to manage building blocks of specialist texts and the terminology of the specialist domain.

As with the case of the paper-based terminology database there is a need to involve domain specialists as possible in the establishment of the electronic database. This work involves many members of *interdisciplinary* teams, including linguists (specifically trained in linguistics), terminology experts (persons with high level of ability in the terminology relevant to a given domain), and computer scientists. All of the above develop a computer tool and method to examine, identify, extract the relevant terminology material from the text corpus. In such cases the terminologist should apply a text-based approach to the terminology. First, the computer should be programmed to scan vast amounts of terminology data in order to automatically identify the terms in the field. The first thing is to establish the key words as in **plough** (alb. *plug*); **cultivator** (alb. *kultivator*); **seeder** (alb. *makinë mbjellese*). At a later stage, the terminologist and the agro-mechanist can work out the syntactic, semantic and pragmatic details of the terms and their contextualized meaning. In so doing the terminologists will be able to create term banks in all aspects of farming and cultivating the soil starting with hand tools such as **picks** *(*alb. *kazmë),* **shovel** *(*alb. *lopatë)* **sickle** *(*alb. *drapër)* , **scythe** (alb. kosë) , to farm implements drawn by animals as in **animal-drawn plough** (alb. *plug i terhequr me kafshe*) to mechanized agricultural machineries such as **combine harvesters** *(*alb. *autokombajnë),* **thresher** *(* alb. *makinë shirëse).* etc. Below there is a representative sample of what an electronic database of any given term should look like.

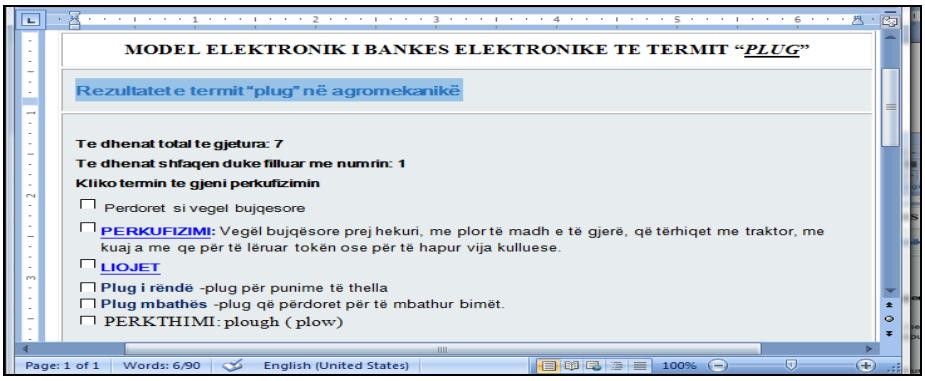

Fig. 2. A sample of an electronic databank in Albanian

4. The importance of terminology database in translation

The translation of specialist documentation is a demanding task on the translator. Not every translator is up to the task. The most challenge lies with the narrow fields of study and knowledge like *agro-mechanics*, *wood-processing*, *genetic engineering*, *sociobiology*, etc., which consist of a very specific terminology .The translation of such specialized texts ask for a full understanding of the lexical inventory of the source and target language, the grammatical structure of the terms, if they are simple terms, two-words terms (*compounds terms*) or expanded terms, that is terms consisting of a series of words. The specialist documentation is vast and involves several major text types like specialist manuals and instruction manuals, learned papers and journals in the field, textbooks, official documentation (*legislation framework*) etc. They are the most difficult ones to nail down appropriately. Most translators lack the appropriate specialist knowledge in the field and fall in the trap of providing approximates in terminology which are reason enough for ambiguity and confusions in communication. Thus terminology databank will allow for equivalent terms circulating both in the source and the target languages.

[11] See Duro as above.

5. Conclusion

The aim of this paper was to present the different approaches in building up the terminology database. It established the need for the work that needs to be undertaken in the specialized domain of knowledge given the conditions of the country and the pressing need to launch such efforts as to join the EU terminology database in the future. The terminology database is a joint effort that requires linguists, terminologists, specialists and experts in the field to work together. A good part focused on how to track terms and a number of features that need to be considered including the grammatical, semantic and pragmatic features.

References

1. http://en.wikipedia.org/wiki/Plough
2. Bhreathnach, U, Cloke. F., Pháidín C. N, "*Terminology for the European Union: The Irish Experience: The GA IATE Project*" Dublin, 2013.
3. Duro, A. '*Terminology as a system*", Panteon, 2001, Tirana.
4. Feka, Th. *"Observations on terminological lexicon in the Albanian language dictionary*", Studies on Lexicon, 1983, Tirana.
5. Kostallari, A. "*Gjuha letrare kombetare shqipe dhe epoka jone*", Tirane, SF, 1984.
6. Merriam Webster dictionary Online.
7. Shumeli, A. "*The formation and functioning of the agromechanics terminology in Albanian as opposed to English*", PHD thesis, Tirana, 2013.
8. Thomai, J. "*Change of semantic meaning in Albanian language*", 1989, Tirana.