

Translating SNOMED CT Terminology into a Minor Language

Olatz Perez-de-Viñaspre and Maite Oronoz

IXA NLP Group

University of the Basque Country UPV/EHU

Donostia

{olatz.perezdevinaspre, maite.oronoz}@ehu.es

Abstract

This paper presents the first attempt to semi-automatically translate SNOMED CT (*Systematized Nomenclature of Medicine – Clinical Terms*) terminology content to Basque, a less resourced language. Thus, it would be possible to build a new clinical healthcare terminology for Basque. We have designed the translation algorithm and the first two phases of the algorithm that feed the SNOMED CT's Terminology content, have been implemented (it is composed of four phases). The goal of the translation is twofold: the enforcement of the use of Basque in the bio-sanitary area and the access to a rich multilingual resource in our language.

1 Introduction

SNOMED Clinical Terms (SNOMED CT) (IHTSDO, 2014) is considered the most comprehensive, multilingual clinical healthcare terminology in the world. The use of a standard clinical terminology improves the quality and health care by enabling consistent representation of meaning in an electronic health record¹.

Osakidetza, the Basque Sanitary System ought to provide its service in the two co-official languages of the Basque Autonomous Community, in Spanish and in Basque. However, and being Basque a minority language in front of the powerful Spanish language, the use of Basque in the documentation services (for example in the Electronic Medical Records (EMR)) of Osakidetza, is almost zero. One of our goals in this work is to offer a medical terminology in Basque to the bio-medical personnel to try to enforce the use of Basque in the bio-sanitary area and in this way protect the

linguistic rights of patients and doctors. Another objective in this work is to be able to access multilingual medical resources in Basque language. To try to reach the mentioned objectives, we want to semi-automatically translate the terminology content of SNOMED CT focusing in some of its main hierarchies.

To achieve our translation goal, we have defined an algorithm that is based on Natural Language Processing (NLP) techniques and that is composed of four phases. In this paper we show the systems and results obtained when developing the first two phases of the algorithm that, in this case, translates English terms into Basque. The first phase of the algorithm is based on the use of multilingual lexical resources, while the second one uses a finite-state approach to obtain Basque equivalent terms using medical affixes and also transcription rules.

In this paper we will leave aside explanations about i) the translation application, ii) the knowledge management and iii) the knowledge representation, and we will focus on term generation. The application framework that manages the terms has been already developed and it is in use. The knowledge representation schema has been designed and implemented and it is also being used (Perez-de-Viñaspre and Oronoz, 2013).

In the rest of the paper after motivating the work and connecting it to other SNOMED CT translations (sections 2 and 3), the algorithm and the material that are needed to implement the first two phases of the translation-algorithm are described (section 4). After that, results are shown and discussed (sections 5 and 6). Finally, some conclusions and future work are listed in the last section (section 7).

2 Background and significance

“Basque is the ancestral language of the Basque people, who inhabit the Basque Country, a region

¹<http://www.ihtsdo.org/snomed-ct/whysnomedct/snomedfeatures/>

spanning an area in northeastern Spain and southwestern France. It is spoken by 27% of Basques in all territories (714,136 out of 2,648,998). Of these, 663,035 live in the Spanish part of the Basque country (Basque Country and Navarre) and the remaining 51,100 live in the French part (Pyrénées-Atlantiques)²". Basque is a minority language in its standardization process and persists between two powerful languages, Spanish and French. Although today Basque holds co-official language status in the Basque Autonomous Community, during centuries Basque was not an official language; it was out of educational systems, out of media, and out of industrial environments. Due to this features, the use of the Basque Language in the bio-sanitary system is low. One of the reasons for translating SNOMED CT is to try to increase the use of the Basque language in this area.

SNOMED CT is a multilingual resource as its concepts are linked to terms in different languages by means of a concept identifier. Thus, terms in our language will be linked to terms in all the languages in which SNOMED CT is released. Besides, as SNOMED CT is part of the Metathesaurus of UMLS (Unified Medical Language System (Bodenreider, 2004)), Basque speakers will have the possibility of accessing other lexical medical resources (RxNorm, MeSH) containing the concepts of SNOMED CT.

SNOMED CT has been already translated to other languages using different techniques. These translations were done either manually (this is the case of the Danish language (Petersen, 2011)), combining automatic translation with manual work (in Chinese, for example (Zhu et al., 2012)), or using exclusively an automatic translation helping system (that is the case of French (Abdoune et al., 2011)). In the design of the translation task, we have followed the guidelines for the translation of SNOMED CT (Høy, 2010) published by the IHTSDO as it is recommended.

3 SNOMED CT

SNOMED CT provides the core terminology for electronic health records and contains more than 296,000 active concepts with their descriptions organized into hierarchies. (Humphreys et al., 1997) shows that SNOMED CT has an acceptable coverage of the terminology needed to record patient

conditions. Concepts are defined by means of description logic axioms and are used also to group terms with the same meaning. Those descriptions are more generally considered as terms.

There are three types of descriptions in SNOMED CT: Fully Specified Names (FSN), Preferred Terms (PT) and Synonyms. Fully Specified Names are the descriptions used to identify the concepts and they usually have a semantic tag in parenthesis that indicates its semantic type and, consequently, its hierarchy. Regarding what we sometimes refer to as "terms" we can distinguish between PTs and Synonyms.

There are 19 hierarchies to organize the content of SNOMED CT (plus 1 hierarchy for meta-data). The concepts of SNOMED CT are grouped into hierarchies as *Clinical finding/disorder*, *Organism*, and so on. For translation purposes it is important to deeply analyze these hierarchies as some of them need to translate all the terms while others as *Organism* only admit the translation of the synonyms (the preferred term should be the taxonomic one). The guidelines for the translation of the hierarchies are given in (Høy, 2010). We want to remark that only the terms classified as PTs and synonyms in SNOMED CT have been taken into consideration for the translation purposes, as the structure (relationships, for example) is the ontological core of SNOMED CT.

Considering the lexical resources available in the bio-sanitary domain for Basque and the SNOMED CT language versions released, two source languages can be used for our translation task: English and Spanish. Basque is classified as a language isolate, and in consequence it is not related to English or Spanish and its linguistic characteristics are far away from both of them. For that reason, no English nor Spanish offers any advantage as translation source. Thus, we deeply analyzed both of them to choose the best option. Our starting point was the Release Format 2 (RF2), Snapshot distributions and the versions dated the 31-07-2012 for English and the 30-10-2012 for Spanish. It must be taken into consideration that the Spanish version of SNOMED CT is a manual translation of the English version.

To choose the source version of SNOMED CT that will be translated, we analyzed aspects as i) general numbers of FSNs, PTs and Synonyms, ii) length of the terms in each language and, ii) the lack of elements in each version. These data help

²http://en.wikipedia.org/wiki/Basque_language (January 23, 2014)

us to come to a decision:

1. The number of active concepts in both languages is the same (296,433) as the Spanish version uses the English concept file. Nevertheless, the number of terms in Spanish is significantly smaller. In Spanish 15,715 concepts lack of PTs and Synonyms.
2. Regarding the length of the PTs and synonyms, we counted the terms containing one token, two tokens, three tokens, four tokens and those with more than four tokens. In the English version the 6.76% of the terms has one token, the 23.28% two and the 20.70% three tokens. That is, quite simple terms compose the half of the synonyms in the lexicon. In the Spanish version, nevertheless, only the 33.79% of the synonyms has three tokens or less, and there are 66.21% synonyms with four tokens or more.

Considering these data, we can conclude that i) the English version is more complete and consistent than the Spanish one, and that ii) the terms in the English version are shorter in length and, in consequence, simpler to translate than the ones in the Spanish version. Thus, we decided to use the English version of SNOMED CT as the translation source as starting point.

We fix the priority between hierarchies for the translation taking into account the number of terms in each hierarchy. The most populated hierarchies are *Clinical finding/disorder* (139,643 concepts) and *Procedure* (75,078 concepts). The next most populated hierarchies are *Organism* (35,870 concepts) and *Body Structure* (26,960). The translation guidelines indicate that the PTs of the organisms should not be translated. For this reason and being conscious of our limitation to translate this huge terminology, we decided to prioritize the translation of the *Clinical finding/disorder*, the *Procedure* and the *Body Structure* hierarchies.

4 Translation Algorithm

We have defined a general algorithm that tries to achieve the translation with an incremental approach. Although the design is general and the algorithm could be used for any language pair, some linguistic resources for the source and objective languages are necessary. In our implementation,

the algorithm takes a term in English as input and obtains one or more equivalent terms in Basque.

The mapping of SNOMED CT with ICD-10 works at concept level. Thus, before executing the implementation of the algorithm the mapping between them should be done (see section 5).

The algorithm is composed of four main phases. The first two phases are already developed and results regarding quantities are given in section 5. The last two phases will be undertaken in the very near future.

We want to remark that all the processes finish in the step numbered as 4 in the algorithm (see Figure 1). The Basque equivalents with their original English terms, and relative information (for instance, the SNOMED CT concept identifier) are stored in an XML document that follows the TermBase eXchange (TBX) (Melby, 2012) international standard (ISO 30042) as exposed in (Perez-de-Viñaspre and Oronoz, 2013). All the lexical resources are stored in another simpler TBX document called ItzulDB (see number 1 in Figure 1). This document is initialized with all the lexical resources available, such as specialized dictionaries and it is enriched with the new translation pairs generated that overcome a confidence threshold with the intention of using them to translate new terms. In this way we achieve feedback.

Let us describe the main phases:

1. *Lexical knowledge*. In this phase of the algorithm (see numbers 1-2-4 in Figure 1), some specialized dictionaries and the English, Spanish and Basque versions of the International Statistical Classification of Diseases and Related Health in its 10th version (ICD-10) are used. ItzulDB is initialized with all the translation pairs (English-Basque) extracted from different dictionaries of the bio-medical domain and the pairs extracted from the ICD-10. For example the input term “abortus” will be stored with all its Basque equivalents “*abortu*”, “*abortatze*” and “*hilaurtze*”. This XML database is enriched with the new elements that are generated when the algorithm is applied (number 4 in Figure 1). Figure 2 shows an example of some translations obtained using *ItzulDB*.
2. *Morphosemantics*. When a simple term (term with a unique token) is not found in *ItzulDB* (number 3 in Figure 1) it is analyzed at word-level, and some generation-rules are used to

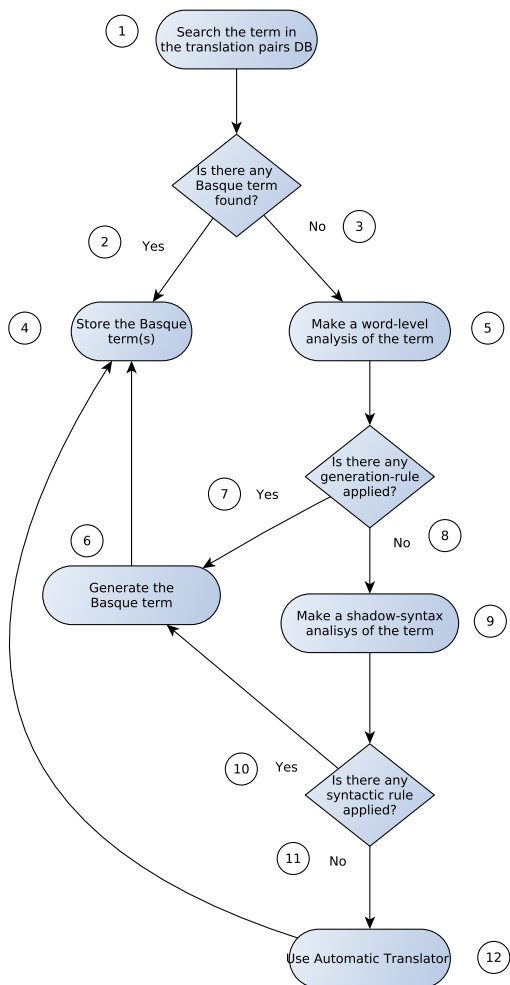


Figure 1: Schema of the Algorithm.

Input term: Deoxyribonucleic acid
Steps in Figure 1 number: 1,2,4
Translation: *Azido desoxirribonukleiko, ADN, DNA*

Figure 2: Terms obtained from *ItzulDB*.

create the translation. We apply medical suffix and prefix equivalences and morphotactic rules, as well as some transcription rules, for this purpose. This is the case in Figure 3.

Input term: Photodermatitis
Steps in Figure 1 number: 3,5,7,6,4
Applied rules:
Identified parts: photo+dermat+itis
Translated parts: foto+dermat+itis
Translation: *Fotodermatitis*

Figure 3: Terms obtained using generation-rules.

3. *Shallow Syntax*. In the case that the input term does not appear in *ItzulDB* and it can not be generated by word-level rules (number 8 in the algorithm), chunk-level generation rules are used. Our hypothesis is that some chunks of the term will appear in *ItzulDB* with their translation. The application should generate the entire term using the translated components (see example in Figure 4).

Input term: Deoxyribonucleic acid sample
Steps in Figure 1 number: 8, 9, 10, 6, 4
Chunks in *ItzulDB*:
1st chunk: Deoxyribonucleic acid
 Basque: *azido desoxirribonukleiko, ADN, DNA*
2nd chunk: sample
 Basque: *lagin*
Translation: *Azido desoxirribonukleikoaren lagin, ADN lagin, DNA lagin*

Figure 4: Terms obtained using chunk-level generation rules.

4. *Machine Translation*. In the last phase, our aim is to use a rule-based automatic translation system called *Matxin* (Mayor et al., 2011) that we want to adapt to the medical domain. Figure 5 shows an attempt of translation with the non adapted translator. For example, *Matxin* translates “colon” as the punctuation mark (“*bi puntu*” or “:”) because it lacks the anatomical meaning.

Input term: Partial excision of oesophagus and interposition of colon
Steps in Figure 1 number: 12, 4
Translation: *Esofagoaren zati baten excisiona eta interpositiona bi puntua*

Figure 5: Terms obtained using *Matxin*.

The IHTSDO organization releases a semi-automatic mapping between SNOMED CT and the ICD-10. By identifying the sense of a concept in SNOMED CT, the best semantic space in the ICD-10 for this concept is searched obtaining linked codes. In this way we can obtain the corresponding Basque term for some of the SNOMED CT concepts through ICD-10. Considering that the structures of SNOMED CT and the ICD-10 are quite different, and that the mapping sometimes has “mapping conditions”, the use of this

resource has been complex, but fruitful for very specialised terms. Although as we said this mapping is the unique source for obtaining very specialised terms, it should be used carefully as the objectives of SNOMED CT and ICD-10 are different. ICD-10 has classification purposes while SNOMED CT has representation purposes.

A brief description of the first two phases of the algorithm is done in the next subsections (subsections 4.1 and 4.2):

4.1 Phase 1: Lexical Resources

The multilingual specialized dictionaries with English and Basque equivalences that have been used to enrich *ItzulDB* in the first phase of the algorithm are:

- *ZT Dictionary*³: This is a dictionary about science and technology that contains areas as medicine, biochemistry, biology... It contains 13,764 English-Basque equivalences.
- *Nursing Dictionary*⁴: It has 5,393 entries in the English-Basque chapter.
- *Glossary of Anatomy*: It contains anatomical terminology (2,578 useful entries) used by University experts in their lectures.
- *ICD-10*⁵: This classification of diseases was translated into Basque in 1996. It is also available in English and in Spanish. The mapping between the different language editions conforming a little dictionary, allowed us to obtain 7,061 equivalences between English and Basque.
- *EuskalTerm*⁶: This terminology bank contains 75,860 entries from which 26,597 term equivalences are labeled as from the biomedical domain.
- *Elhuyar Dictionary*⁷: This English-Basque dictionary, is a general dictionary that contains 39,164 equivalences from English to Basque.

All these quite different dictionaries have been preprocessed in order to initialize *ItzulDB*. *Elhuyar Dictionary* is a general dictionary that has

³<http://zthiztegia.elhuyar.org>

⁴<http://www.ehu.es/euskalosasuna/Erizaintza2.pdf>

⁵<http://www.ehu.es/PAT/Glosarios/GNS10.txt>

⁶<http://www.euskadi.net/euskalterm>

⁷<http://hiztegiak.elhuyar.org/en>

both not domains pairs but also contains some specialized terminology. This general dictionary will help i) in the translation of not domain terms and ii) also in the translation of the chunks in Phase 3, and thus, on the generation of new terms in Basque.

4.2 Phase 2: Finite State Transducers and Biomedical Affixes

A first approach to this work is presented in (Perez-de-Viñaspre et al., 2013). In that work, finite state transducers described in Foma (Hulden, 2009) are used to automatically identify the affixes in English Medical terms and by means of affix translation pairs, to generate the equivalent terms in Basque. We observed that the behavior of the roots in this type of words is similar to prefixes, so, we will not make distinction between them and we will name them prefixes. A list of 826 prefixes and 143 suffixes with medical meanings was manually translated. An evaluation of the system was performed in a Gold Standard of 885 English-Basque pairs. The Gold Standard was composed of the simple terms that were previously translated in the first phase of the algorithm. A precision of 93% and a recall of 41% were obtained.

In that occasion, only SNOMED CT terms for which all the prefixes and suffixes were identified were translated. For example, terms with the prefix “phat” were not translated as this affix does not appear in the prefixes and suffixes list. For instance, the “hypophosphatemia” term was not translated even though the “hypo”, “phos” and “emia” affixes were identified.

We have improved this work by increasing the number of affixes and implementing transcription rules from English/Latin/Greek to Basque.

Figure 6 will help us to get a wider view of the work exposed. The input term “symphysiolysis” is split into the possible affix combination in the first step (“sym+physio+lysis” or “sym+physi+o+lysis”). Then, those affixes are translated by means of its equivalents in Basque (“sim+fisi+lisi” or “sim+fisi+o+lisi”). And finally, by means of morphotactic rules, the well-formed Basque term is composed (in both cases “sinfisiolisi” is generated).

5 Results

Considering the huge size of the descriptions in SNOMED CT and to make the translation pro-

Table 1: Results of the translation.

	Disorder		Finding		Body Structure		Procedure	
	#Synonyms	#Matches	#Synonyms	#Matches	#Synonyms	#Matches	#Synonyms	#Matches
ICD-10 mapping	11,227	-	1,878	-	0	-	0	-
In dictionaries	4,804	3,488	1,836	915	5,896	2,992	778	473
ZT Dictionary	1,104	883	367	311	1,812	1,212	293	253
Nursing Dictionary	437	350	340	245	978	725	199	157
Glossary of Anatomy	3	3	10	8	1,982	1,431	2	2
ICD-10	2,434	2,308	216	195	410	370	5	4
EuskalTerm	906	596	442	306	2,346	1,423	202	155
Elhuyar	299	135	956	300	1,090	367	270	91
Morphosemantics	2,620	2,184	705	578	970	779	1,551	1,362
Total	17,627	5,672	4,419	1,493	6,866	3,771	2,329	1,835

Input term: symphysiolysis
Identified affixes: sym+physio+lysis, sym+physi+o+lysis
Translation of the affixes: sim+fisio+lisi, sim+fisi+o+lisi
Morphotactics output term: <i>sinfisiolisi</i>

Figure 6: Term translated by means of affix equivalences.

cess easy to handle, we have divided it into hierarchies. The *Clinical finding/disorder* hierarchy is specially populated so we have split it considering its semantic tags: *disorders* and *findings*. In addition, the terms from the *Procedure* and *Body Structure* hierarchies have been evaluated too.

Before showing the results, we want to remark some aspects of the evaluation:

- Phase 1: the evaluation has been performed in terms of *quantity*, not of *quality* of the equivalent terms obtained. As the used resources are dictionaries manually generated by lexicographers and domain experts, the quality of the Basque terms is assumed. In any case, and due to the fact that Basque is in its standardization process, the orthographic correctness of the descriptions (see section 6) will be manually checked in the near future.
- Phase 2: the quality of the generated terms could be measured extrapolating the results in the evaluation of the baseline system described in subsection 4.2. That is, 93% precision and 41% recall. The quantity results are shown considering the improvements described in the same subsection.

Table 1 shows the results for the mentioned hierarchies and semantic tags when the translation is

performed using both methods: dictionary matching and morphosemantics. Remind that in a previous phase a concept level mapping is completed between SNOMED CT and ICD-10. The first row in Table 1 labeled as “ICD-10 mapping” shows that it is relevant only for the *Clinical disorders and findings* hierarchy, being the *disorder* semantic tag the most benefited one with 11,228 equivalences. The remainder of the results is given at term level.

We made a distinction between the number of obtained Basque terms (1st column, labeled as “#Synonyms”) and the number of English terms translated (2nd column, labeled as “#Matches”). Let us see the difference between those two columns looking at the numbers in Table 1. For example, in the *disorder* semantic tag there are 3,488 matches (3,488 original English terms translated), but the number of obtained Basque terms is 4,804 (adding the number of equivalents of all the dictionaries). The reason is that the same input term may have synonyms or even the same equivalent term given by different dictionaries. For example, for the term “allopathy”, the same term “alopatia” is obtained in the ZT and Nursing dictionaries (this equivalence will be counted in both ZT and Nursing dictionaries rows).

Table 2 shows the number of tokens in the original English terms. This table refers not to the concepts, but to the terms in the source SNOMED CT in English. The first row shows the number of English terms to which we obtained a Basque equivalent or synonym, the second one the total of English terms and finally, the last row the percentage of translated terms.

Table 3 gives the overall numbers of the translated concepts, in order to take a wide view of the process done.

Let us see the highlights of the results for each

Table 2: Results of the translation regarding the number of tokens of the original term.

		1 token	2 tokens	3 tokens	4 tokens	> 4 tokens	Total
Disorder	Translated Terms	3,315	1,114	538	279	426	5,672
	Terms in total	4,066	22,023	24,036	20,005	37,316	107,446
	Percentage	81.53%	5.06%	2.24%	1.40%	1.14%	5.27%
Finding	Translated Terms	1,222	158	39	20	54	1,493
	Terms in total	1,830	8,837	10,980	9,814	19,106	50,567
	Percentage	66.78%	1.79%	0.36%	0.20%	0.28%	2.95%
Body Structure	Translated Terms	1,942	1,416	334	66	13	3,771
	Terms in total	2,692	11,519	12,575	10,903	21,631	59,320
	Percentage	72.14%	12.29%	2.66%	0.61%	0.06%	6.36%
Procedure	Translated Terms	1,741	80	11	2	1	1,835
	Terms in total	1,982	9,966	15,848	16,578	37,695	82,069
	Percentage	87.84%	0.80%	0.07%	0.01%	0.003%	2.24%

Table 3: Overall results.

	Disorder	Finding	Body Structure	Procedure
Translated Concepts	14,125	2,777	3,231	1,502
Concepts in total	65,386	33,204	31,105	82,069
Percentage	21.60%	8.36%	10.39%	1.83%

hierarchy or semantic tag:

- 21.60% of the *disorders* has been translated (see Table 3). This can be considered a very good result. The ICD-10 mapping produces the majority of the translations as it could be expected in this hierarchy (11,227 synonyms obtained). In Table 2 the strength of the morphosemantics phase is evident as the 81.53% of the simple terms is translated.
- The *finding* semantic tag is the most balanced, as no one of the algorithm phase’s contribution outlines. The translation of the 8.36% of the concepts is achieved.
- Regarding the results of the *Body Structure* hierarchy, Table 1 shows that the Glossary of Anatomy only contributes in this area. The 10.39% of the concepts get a Basque equivalent.
- In the translation of the *Procedure* hierarchy the dictionaries do not help much as shown in Table 1. In contrast, the morphosemantics contribution allows to translate the 87.84% of the simple terms (see Table 2).

6 Discussion

Some general dictionaries as the ZT dictionary usually contribute in the translation of most of the terms, while more specialized dictionaries only provide translations in the terms related to their

domain. For example, both dictionaries, the ZT dictionary and the Nursing dictionary, obtained the Basque terms “mikrozefalia” for “microcephaly” and “metatartso” for “metatarsus”. The ICD-10 mapping contributed mainly in the translation of the disorders, and the Glossary of Anatomy in the translation of terms from the Body Structure hierarchy. Sometimes more than an equivalent in Basque is obtained in the translation. For example, for the term “leprosy” we got the equivalents “legen beltz”, “legen” and “legendar”. Some problems were detected in the Basque terms regarding the standard orthography (the ICD-10 was translated in 1996 and the spelling rules have changed since then) and the form of the word (some obtain the word in finite forms, i.e. “abdomena” for “abdomen” and other in non finite form, “abdomen”).

To which the terms generated by finite-state transducers concern, we detected many new affixes from the SNOMED CT terms that do not appear in our lexicon. Even most of those affixes will be correctly transcribed by our transducers, experts insist on enriching the lexicon with new pairs.

7 Conclusions

We have designed a translation algorithm for the multilingual terminology content of SNOMED CT and we have implemented the first two phases. On the one hand, lexical resources feed our database, and on the other hand, Basque equivalents are generated using transducers and medical and biologi-

cal affixes.

Dictionaries provide Basque equivalents of any term length (i.e. unique and multitoken terms) while transducers get as input unique token terms.

In both translation methods results for the most populated hierarchies are shown even though they are applied for all the hierarchies in SNOMED CT. When using lexical resources, results are promising and the contribution of the ICD-10 mapping is remarkable. We obtained the equivalents in Basque of 21.60% of the disorders.

In any case, as we said before, our objective in the future is that specialist in medical terminology can check the quality of the obtained terms and correct them with the help of a domain corpus in Basque. A platform is being developed for this purpose. After the evaluation, and only if it reaches high quality results, our aim is to contact SNOMED CT providers to offer them the result of our work, that at the moment only pertains to the research area.

Regarding the developed systems evaluation, the system used in the first phase extracts English-Basque pairs from dictionaries, so being quite a simple system, does not need of a deep evaluation. A first evaluation of the system that generates terms using medical affixes has been presented. At present, we are evaluating the improvements of this second system with promising results.

In a near future, we want to implement the remainder of the phases in the algorithm: the use of syntax rules for term generation, and the adaptation of the machine translation tool. The promising results in this first approximation encourage us in the way to semi-automatically generate a version in Basque of SNOMED CT.

Acknowledgments

The authors would like to thank Mikel Lersundi for his help. This work was partially supported by the European Commission (325099), the Spanish Ministry of Science and Innovation (TIN2012-38584-C06-02) and the Basque Government (IT344-10 and IE12-333). Olatz Perez-de-Viñaspre's work is funded by a PhD grant from the Basque Government (BFI-2011-389).

References

Hocine Abdoune, Tayeb Merabti, Stéfan J. Darmoni, and Michel Joubert. 2011. Assisting the Translation of the CORE Subset of SNOMED CT Into French.

In Anne Moen, Stig Kjær Andersen, Jos Aarts, and Petter Hurlen, editors, *Studies in Health Technology and Informatics*, volume 169, pages 819–823.

Olivier Bodenreider. 2004. The Unified Medical Language System (UMLS): Integrating Biomedical Terminology. *Nucleic acids research*, 32(suppl 1):D267–D270.

Asta Høy. 2010. Guidelines for Translation of SNOMED CT. Technical Report version 2.0, International Health Terminology Standards Development Organization IHTSDO.

M. Hulden. 2009. Foma: a Finite-State Compiler and Library. In *Proceedings of EACL 2009*, pages 29–32, Stroudsburg, PA, USA.

Betsy L Humphreys, Alexa T McCray, and May L Cheh. 1997. Evaluating the coverage of controlled health data terminologies: report on the results of the NLM/AHCPR large scale vocabulary test. *Journal of the American Medical Informatics Association*, 4(6):484–500.

International Health Terminology Standards Development Organisation IHTSDO. 2014. SNOMED CT Starter Guide. February 2014. Technical report, International Health Terminology Standards Development Organisation.

Aingeru Mayor, Iñaki Alegria, Arantza Diaz de Ilaraza, Gorka Labaka, Mikel Lersundi, and Kepa Sarasola. 2011. Matxin, an Open-source Rule-based Machine Translation System for Basque. *Machine Translation*, 25:53–82. 10.1007/s10590-011-9092-y.

Alan K. Melby. 2012. Terminology in the Age of Multilingual Corpora. *The Journal of Specialised Translation*, 18:7–29, July.

Olatz Perez-de-Viñaspre and Maite Oronoz. 2013. An XML Based TBX Framework to Represent Multilingual SNOMED CT for Translation. In *Advances in Artificial Intelligence and Its Applications*, pages 419–429. Springer.

Olatz Perez-de-Viñaspre, Maite Oronoz, Manex Agirrezabal, and Mikel Lersundi. 2013. A Finite-State Approach to Translate SNOMED CT Terms into Basque Using Medical Prefixes and Suffixes. *Finite State Methods and Natural Language Processing*, page 99.

Palle G. Petersen. 2011. How to Manage the Translation of a Terminology. Presentation at the IHTSDO October 2011 Conference and Showcase, October.

Yanhui Zhu, Huiting Pan, Lei Zhou, Wei Zhao, Ana Chen, Ulrich Andersen, Shuxiang Pan, Lixin Tian, and Jianbo Lei. 2012. Translation and Localization of SNOMED CT in China: A pilot study. *Artificial Intelligence in Medicine*, 54(2):147–149.