KabiTermICD: Nested Term Based Translation of the ICD-10-CM into a Minor Language

Olatz Perez-de-Viñaspre, Maite Oronoz, Natalia Elvira
IXA NLP Group (UPV/EHU), Osakidetza
olatz.perezdevinaspre@ehu.eus, maite.oronoz@ehu.eus, NATALIA.ELVIRAGARCIA@osakidetza.eus

Abstract

In this paper a system called KabiTermICD is presented, which automatically translates ICD-10-CM English terms into a less resourced language, Basque. The lack of a big enough specialized bilingual corpus between Basque and English inspired the creation of KabiTermICD system for the translation of medical terminology. This system is based on the semantic structures that complex terms have, where medical terms are nested into longer terms. The technology used is based on finite state transducers and previously developed multilingual medical lexicons (SNOMED CT, for instance). The results, showed a big time saving for experts to translate ICD-10-CM as around the 60% of the translated descriptions did not need to be post-edited, and most of the remaining needed a smaller post-editing effort.

Keywords: Medical term automatic translation, Finite state transducers, ICD-10-CM, Coding standards

1. Introduction and Background

In this paper KabiTermICD is presented, a system for the automatic translation of the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) from English into Basque.

The main objective of this work is to lighten the translation work of lexicographers by offering them automatically generated accurate terms in Basque. Instead of working on the translations, these professionals would review the automatically generated ones and if necessary, post-edit or correct them.

Basque is an isolated language spoken by more than 700.000 people in the Basque Country. It has not related language, and it coexists with two big languages: Spanish and French. Basque is one of the two official languages in the Basque Autonomous Community, and thus, all the official documents are published both in Spanish and Basque. By law, the Basque Sanitary System called Osakidetza should have its documentation in both languages too. Documents related to the administration are bilingual in Osakidetza but health reports are to be managed distinctively due to our context: Osakidetza has a centralized system where each doctor can have access to all the medical reports about their patients written by different services and health professionals. Not all the doctors in Osakidetza speak Basque language and, as a consequence, to safeguard patient's medical security, electronic medical records are written only in Spanish. In the future, the idea of Osakidetza is to have Electronic Health Records (EHRs) in both languages, among others, to protect patients and doctors linguistic rights while guaranteeing security. We are working in the same direction and collaborating with Osakidetza.

In order to go ahead on with the linguistic normalization of Basque inside the health system (also called language planning), works on establishing Basque medical terminology have been done in the last years. The Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) was already automatically translated from English into Basque (Perez-de Viñaspre and Oronoz, 2014; Perez-de Viñaspre and Oronoz, 2015). Different techniques were used to translate the terminological content of SNOMED CT: 1) the reuse of already developed linguistic resources, 2) finite state transducers[1] that use biomedical affixes to analyze, split, translate and build clinical terms (Perez-de-Viñaspre et al., 2013), 3) transducers that take advantage of medical terms nested into longer terms and, finally 4) the adaptation of a rule-based machine translation system, Matxin, to the medical domain, obtaining MatxinMed. The system developed to implement the third step is called KabiTerm (*Kabi* means "nested" in Basque in reference to the nested terms used for the translation). KabiTermICD is the adaptation of KabiTerm in order to translate the ICD-10-CM classification.

ICD-10 and SNOMED CT have different purposes and this reflects in the construction of some terms. ICD-10 is a very useful classification for statistical recording but the rich semantic structure of SNOMED CT makes it very valuable and adds meaning to the EHRs, besides of being adequate to detail for clinical recording.

The ICD-10, owned and published by the World Health Classification (WHO), replaced in 1999 the previous ICD-9 version with the purpose of coding and classifying mortality data from death certificates[2]. According to WHO, ICD is used by all WHO members (194 Member States), has been translated into 43 languages and in most countries (117) is used to report mortality data, a primary indicator of health status[3]. The clinical modification (ICD-10-CM) improves ICD-10 by the addition of relevant information to ambulatory and managed care encounters; by expanding injury codes; by combinations of diagnosis/symptom codes

[1]"A finite-state transducer (FST) is a finite-state machine with two memory tapes, following the terminology for Turing machines: an input tape and an output tape". In Wikipedia. Retrieved March 5, 2018, from https://en.wikipedia.org/wiki/Finite-state_transducer

[2]https://www.cdc.gov/nchs/icd/icd10cm.htm

[3]http://www.who.int/classifications/icd/en/

to reduce the number of codes to describe a condition, etc. It also increments the number of digits in the codes to allow greater coding specificity (from 3 to 5 digits or from 3 to 7 digits). Each diagnostic term in ICD is linked to one or more numeric codes.

In Spain, by law, from the 1st of January of 2016, all the hospitals must code their diagnostic terms with the CIE-10-ES classification. In fact, this is a translation of the ICD-10-CM classification created and validated by the Spanish Ministry of Health, Social Services and Equality[4]. In 1996 the Health Department of the Basque Autonomous Community translated into Basque the CIE-10 classification, obtaining the GNS-10 classification (GNS is the name in Basque for ICD). Nowadays, the expansion related to the clinical modification is needed, and we are working with Osakidetza, using KabiTermICD to obtain the terms in this extension. The amount of translations needed is still big, as the ICD-10-CM is composed of 93,830 codes, and the ICD-10 in Basque of 12,619. The 1996's ICD-10 version is not updated to the last Basque orthographic rules. In this work, ICD-10-CM is being translated into Basque to obtain the GNS-10-MK translation.

Figure 1 shows part of the ICD-10-CM classification. More specifically, the codes for "viral pneumonia". As we said before, being a classification, the medical terms appearing in it have some characteristics. For example, it gathers terms as "Viral pneumonia, **not elsewhere classified**" (main code J12), "**Other** viral pneumonia" (more specific code J12.8) or "Viral pneumonia, **unspecified**" (code 12.9). The texts "not elsewhere classified", "other" or "unspecified" do not usually appear in, for example, SNOMED CT.

J12 **Viral pneumonia, not elsewhere classified**

Incl.: bronchopneumonia due to viruses other than influenza viruses

Excl.: congenital rubella pneumonitis (P35.0)

pneumonia:
- aspiration (due to):
  - NOS (J69.0)
  - anaesthesia during:
    - labour and delivery (O74.0)
    - pregnancy (O29.0)
    - puerperium (O89.0)
  - neonatal (P24.9)
  - solids and liquids (J69.-)
- in influenza (J09, J10.0, J11.0)
- interstitial NOS (J84.9)
- lipid (J69.1)
- viral, congenital (P23.0)

severe acute respiratory syndrome [SARS] (U04.9)

J12.0 **Adenoviral pneumonia**

J12.1 **Respiratory syncytial virus pneumonia**

J12.2 **Parainfluenza virus pneumonia**

J12.3 **Human metapneumovirus pneumonia**

J12.8 **Other viral pneumonia**

J12.9 **Viral pneumonia, unspecified**

Figure 1: Example of classification of the diagnostic term 'Viral Pneumonia'.

Langlais et al. (2008) define two methods to translate terminology automatically. The "generative" methods where new target words are generated from previously unseen source words (human expertise or machine learning techniques can be used). And, non-generative methods where word translations can be found in parallel corpora (word-alignment methods). As there are not English-Basque bilingual medical corpora, the methods used to translate medical terms from English into Basque are all generative.

Some works about the translation of the Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) into different languages have been performed using different techniques: i) The translation into French was done using exclusively automatic translation helping systems (Abdoune et al., 2011), ii) for Chinese, automatic translation and manual work were combined (Zhu et al., 2012), iii) the Danish translation was obtained manually (Petersen, 2011) and iv) three kinds of translations were used to translate SNOMED CT into German (Schulz et al., 2013).

However no work was found in the automatic translation of the ICD medical classification into any language, and much less, into a minority language as Basque, which does not have parallel corpora.

2. KabiTermICD

KabiTermICD is an extension of the KabiTerm system. KabiTerm is based in finite state transducers to obtain Basque equivalents of English terms and it has been implemented using the Foma library and compiler (Hulden, 2009). This system is adequate for the translation in language pairs lacking of bilingual corpus.

The current version of KabiTerm takes a complex English term and, providing the resources available, proposes Basque equivalents. It uses terms that appear within complex terms to translate those complex terms into the Basque language. Here, resources are understood to mean the English-Basque equivalents on the one hand, and the translation patterns on the other. As explained later on in this section, in order to facilitate the work carried out by KabiTerm, an analyser called AnaMed has been developed. This analyser is responsible for obtaining the information required by KabiTerm. It also identifies and prepares the nested terms, leaving KabiTerm free to focus solely on the translation into Basque.

2.1. AnaMed: Medical Term Analyser

In addition to the analysis of linguistic information, AnaMed also identifies SNOMED CT terms and eponyms in a given text. Eponyms are proper nouns that appear in the designation of certain concepts. AnaMed has been developed for English and Basque, and the aim is to adapt it also to the Spanish language in the future.

Initially, only the English version of the AnaMed analyser was developed, in response to our need for a tool to search for the information required by the KabiTerm system outlined in this section. In other words, the information gathered by AnaMed is information that may prove necessary for automatic machine translation from English into Basque. However, since AnaMed can easily be adapted to other languages, it was decided to develop a Basque-language version, since this might prove useful for the drafting of medical reports in Basque.

[4]https://www.msssi.gob.es/en/home.htm

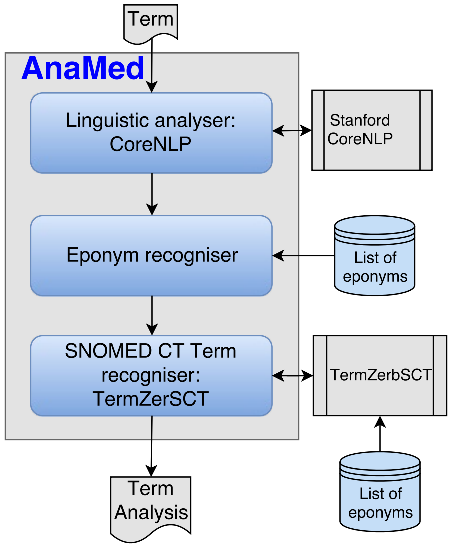

Figure 2: AnaMed analyser architecture.

AnaMed is based on an automatic analysis system and integrates the identification of both eponyms and SNOMED CT terms. The architecture of AnaMed is shown in Figure 2. The following is an explanation of the analysis process using a figure.

- **Linguistic analyser, CoreNLP:** The Stanford CoreNLP tool (Manning et al., 2014) was used as the starting point for the development of the English analyser, along with the Python wrapper for Stanford CoreNLP, developed by Dustin Smith[5]. The tokeniser, the morphological analyser and the tagger from the linguistic analyser were used to identify tokens' lemmas and parts of speech. In addition to this information, token offsets and, named entity tags, were also integrated into the analyser.
- **Eponym recogniser:** By adding a second module the analyser was given eponym identification capability. Eponyms are very common in medical terminology, particularly in the names of diseases and syndromes. The terms "Down syndrome" and "Alzheimer's disease" are good examples of this.
- **SNOMED CT term recogniser:** Finally, the SNOMED CT term identifier was added to the AnaMed analyser. We adapted the TermZerSCT terminology server to identify SNOMED CT terms. SNOMED CT contains a vast amount of terminology (around 300,000 concepts) which takes time to process. TermZerSCT enables faster terminology content management, and when the server is running the reception of information about SNOMED CT terms is almost instantaneous, with a minimum waiting period.

  Using the TermZerSCT server, AnaMed identifies the nested terms located within complex terms, enabling us to analyse the structure of said complex terms. Moreover, it also groups nested terms together using underscores ("_"). For example, in the complex term "unstable diabetes mellitus" it identifies two nested terms: the qualifier "unstable" and the disorder "diabetes mellitus". Thanks to this identification, in addition to providing the complete analysis, AnaMed also gives us the structure (QUALIFIER+DISORDER) and the grouping ("unstable diabetes_mellitus"), information which is extremely useful for KabiTerm.

[5]https://github.com/dasmith/stanford-corenlp-python (accesed May 9, 2017)

2.2. KabiTerm's transducers

KabiTerm's operating process is shown in Figure 3:

1. First of all, AnaMed analyses the input term, identifying and grouping any nested term contained within it. In the case of the term "malignant neoplasm of small intestine, unspecified", "malignant_neoplasm" is a complex term and "small" and "intestine" are simple terms.

2. Secondly, the system calls the transducer responsible for identifying the Basque translation patterns and tagging the nested terms. Thus, this transducer applies the appropriate Basque translation rule and it attaches the tags required to translate the nested terms into Basque. That is to say, the translation rules are applied by means of new tags that imply information to generate the Basque equivalent. In such case, the transducer identifies the structure DISORDER+of+QUALIFIER+BODYSTRUCTURE and applies the corresponding Basque translation rule. It tags "malignant_neoplasm" with "—DIS+a" because it is a disorder and it appends the singular article mark ("+a") required in Basque, it tags "small as a qualifier ("—QUA") and it tags "intestine" with "—BOD+areM" as in addition to being a body part, the term also requires a declension in Basque ("+areM" in this particular case). Besides, the transducer also adds a change of order tag, indicating that the last two terms (excluding everything after the coma) should be moved to the beginning ("&AzkenBiakLehenera"). At this point, there are 72 translation patterns defined.

3. The next step involves rearranging the nested term, following the instructions provided in the tag added during the previous step ("&AzkenBiakLehenera"). Thus, "small" and "intestine" are moved to the beginning as this is the correct place for them in Basque.

4. In the fourth step the system calls up the transducer responsible for translating nested terms into Basque. This transducer provides us with two Basque equivalent terms: *"txiki&&&ADJK heste+areM neoplasia_gaizto+a , zehaztugabea"* (the semantic tags disappear and the output is the Basque equivalent of each English term).

5. Next, rearrangement of adjectives and determination of plurals is executed. There are two kinds of adjectives in the Basque language, those that go after the noun (*izenondo*) and those that go before it (*izenlagun*). In this example, *"txiki"* is an *izenondo* and since

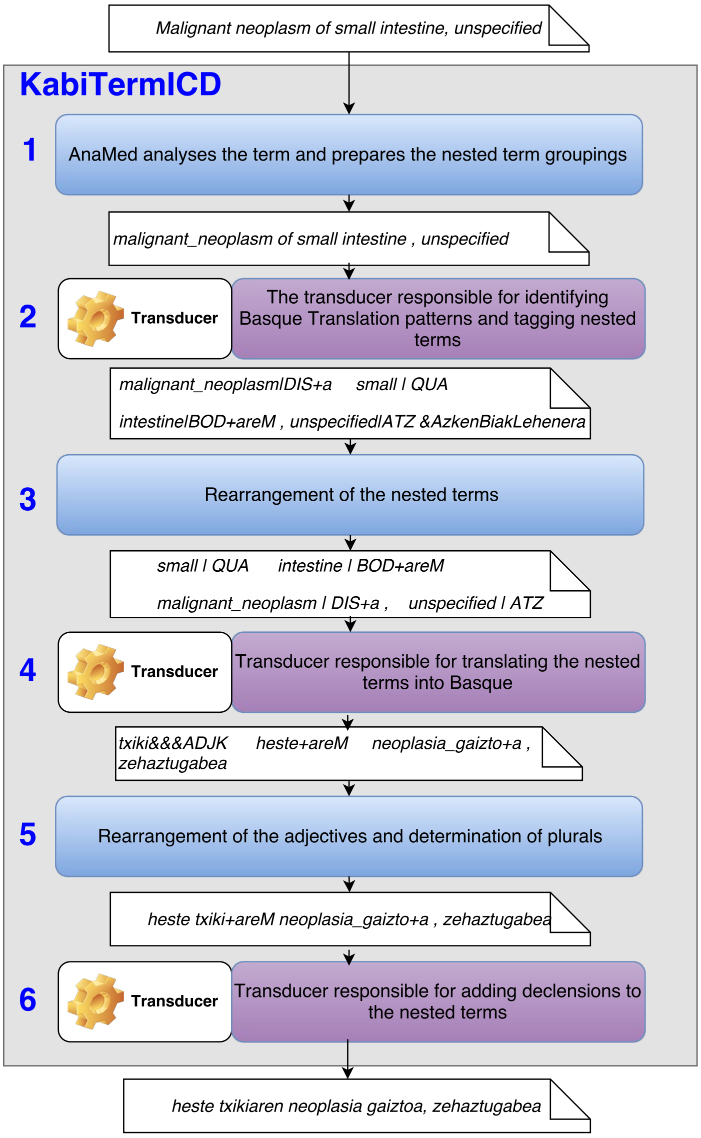

Figure 3: Examples of KabiTermICD's architecture and functioning.

it must go after the name (the &&&ADJK mark expresses that information), said adjective is rearranged after the following term (*"heste"*, translation of "intestine"). Even though this is not the case in the example used here, if the terms contain a nested term in plural in the original English term, in this step its declensions are updated to reflect its plural status.

6. Finally, a new transducer is called up to add the declensions to the nested terms, and thus, it obtains the compound Basque description *"heste txikiaren neoplasia gaiztoa, zehaztugabea"*.

2.3. Translation of ICD-10-CM

As mentioned before, KabiTerm was originally created to translate SNOMED CT's terms into Basque. The characteristics that the descriptions of ICD-10-CM have regarding classification needs, made us include 16 new rules describing the most general structures by now. Being an ongoing process, those structures are still being set and the inclusion of more rules is foreseen in the near future.

The translation of ICD-10-CM was designed as an incremental process that takes advantage of the already translated terms and checked descriptions to generate new ones (Figure 4). For instance, having "malignant" and "neoplasm" terms equivalents in Basque (*"gaizto"* and *"neoplasia"* respectively) KabiTermICD is able to translate "malignant neoplasm" into Basque as *"neoplasia gaizto"*. Similarly, from "lower" and "lip" KabiTermICD generates *"beheko ezpain"* and *"ezpain azpiko"* in Basque. In this case, the expert took *"beheko ezpain"* as the valid translation. This new equivalences may be useful to translate more complex terms, and therefore, the lexicons are enriched with this new equivalences, and the transducers are recompiled. In a second execution of the incremental algorithm (see the second colummn in Figure 4) using the "malignant neoplasm" and "lower lip" nested terms, KabiTermICD is able to translate "malignant neoplasm of lower lip" as *"beheko ezpainaren neoplasia gaizto"*, and in this case the expert post-edited the translation to improve it (*"ezpaineko"* instead of *"ezpainaren"*). Once this new term is included in the lexicons and the transducers recompiled, in subsequent applications of the algorithm KabiTermICD translates terms as “malignant neoplasm of lower lip, inner aspect” as shown in Figure 5.

At this point, there are around 100,000 terms included in the lexicons, divided by their corresponding hierarchy, such as disorder, or body structure. Those lexicons were initialized with the term-equivalent pairs generated during the translation of SNOMED CT, and are extended with the new pairs generated during the translation of ICD-10-CM.

As there is not a Gold Standard so the translated descriptions could be evaluated, an expert is validating one by one all the descriptions to create an ICD-10-CM version in Basque. The work was evaluated by comparing the validated descriptions with the ones created automatically. That is to say, the chosen Basque translation is checked whereas it has been automatically generated, or otherwise it has been post-edited by the expert. Validation means accepting or post-editing a description in a process.

For the above mentioned validation, the web page shown in Figure 5 was developed. This web page shows a list of codes that have been translated automatically. Each of the codes is linked with its corresponding hierarchy inside ICD-10-CM offered by the National Cancer Institute’s Enterprise Vocabulary Services[6].

In addition to the information regarding the code, the corresponding descriptions in English and Spanish (“Malignant neoplasm of lower lip, inner aspect” and *“Neoplasia maligna de labio inferior, cara interna”* respectively) are also shown, as reference for the validation. Even if the source is the English description, the sociolinguistics context makes Spanish a compulsory reference for the validation. Finally, all the Basque candidates are disclosed so the expert can choose one single candidate as the referenced one (in this case *“beheko ezpaineko neoplasia gaiztoa, barrualdea”*). If any of the candidates is good enough, the expert can edit the text to get the good translation. The web page also allows the expert to let some of the descriptions to be reviewed later.

The validation, and so, the translation, is being made by phases. In the first phase, the translations obtained from dictionaries were loaded, and so a first set of around 6,000 codes was given to the expert. For example, the ICD-10-CM term ”Tuberculosis of lung” corresponding to the code A150, was translated by means of the Basque Terminology Bank called Euskalterm, and given its Basque equivalent *”arnas tuberkulosi”*. Once this set of descriptions was corrected, the lexicons of the transducers were updated obtaining the transducers with the corrections included. In this case, the lexicons were not extended but corrected. It is worth to remember, that the lexicons were initially loaded by the automatic translation of SNOMED CT, and so, around 85,000 term-equivalent pairs were already loaded on those lexicons (excluding synonyms). Even if they were included in the initial lexicons, they were not validated by experts, so they need to be uploaded with the corrections made by experts.

In the following phases, the translations obtained by KabiTermICD are loaded to the validation interface. In each phase, a new set of validated translation pairs is included into the lexicons, and in addition, new rules that describe ICD-10-CM descriptions are written to improve the recall of our system.

At this point, 16,512 descriptions have been translated out of 93,830, and from them, 11,104 have been already validated by the expert.

[6]https://ncitERMS.nci.nih.gov/ncitbrowser/pages/hierarchy.jsf

3. Results and discussion

In this section, the focus will be set on the translations validated by the expert. The expert is a translator employed by Osakidetza with an extend background on medical terminology management.

From the 11,104 validated descriptions available nowadays, 5,663 were initially translated by KabiTermICD, and that is the set used to calculate the results.

Around 60% of the descriptions were accepted without any kind of post-editing work (3,379 descriptions). Regarding the remaining 40% with post-editing needs, Table 1 shows the edit distance between the corrected description and the source translation. Edit distance is a measure to quantify the similarity between two words. For that purpose, edit distance counts the minimum number of operations required to transform one word into the other. Even if there are different definitions for the operations, the most common metric is the Levenshtein Distance (Levenshtein, 1966), and the operations are the removal, insertion and substitution of a single character. The edit distances showed in this work were measured according Levenshtein’s definition.

Edit distance	Quantity	Percentage
1	109	4.77
2	39	1.71
3	216	9.46
4	986	43.17
5	74	3.24
6	59	2.58
7	63	2.76
8	113	4.95
9	71	3.11
10 - 14	225	9.85
15 - 19	198	8.67
20 - 24	67	2.93
25 - 29	38	1.66
30 - 34	13	0.57
35 - 39	9	0.39
40 - 44	3	0.13
45	1	0.04
Total	2,284	100

Table 1: Edit distance of the post-edited descriptions

After analysing the corrections made to the source translations, correlation between the edit distance and the correction type was found out. In the following lines, the main correction types found on them are listed.

- Edit distance 1: two main mistakes were corrected, spelling errors (row 1 in Table 2) and overproduction

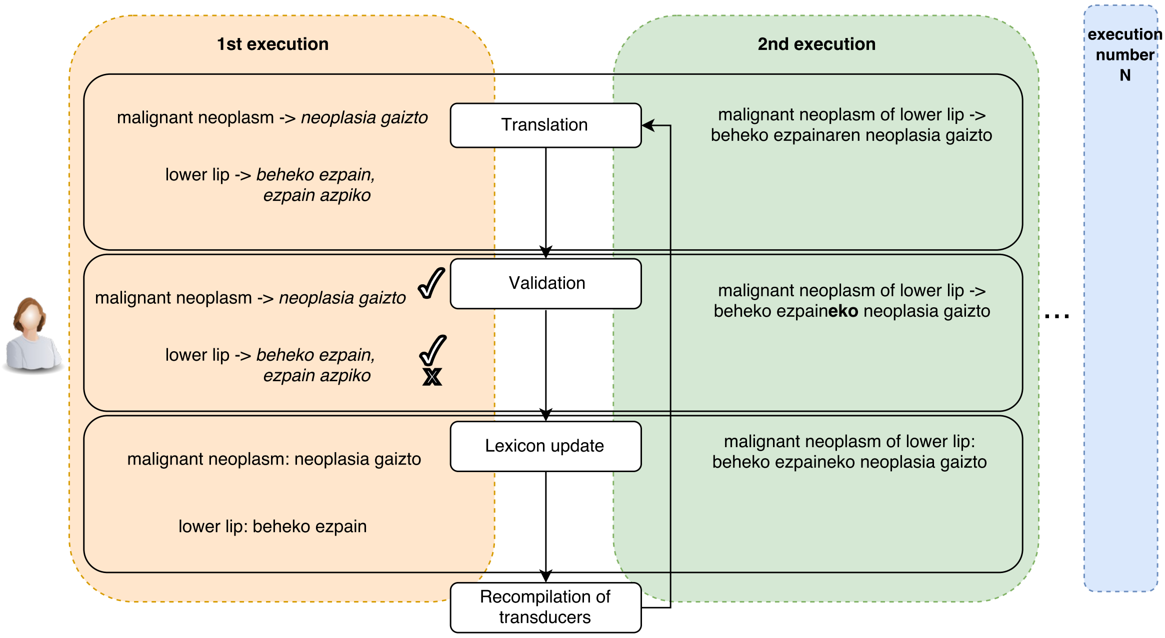

Figure 4: Diagram of the incremental functioning of the translation process.

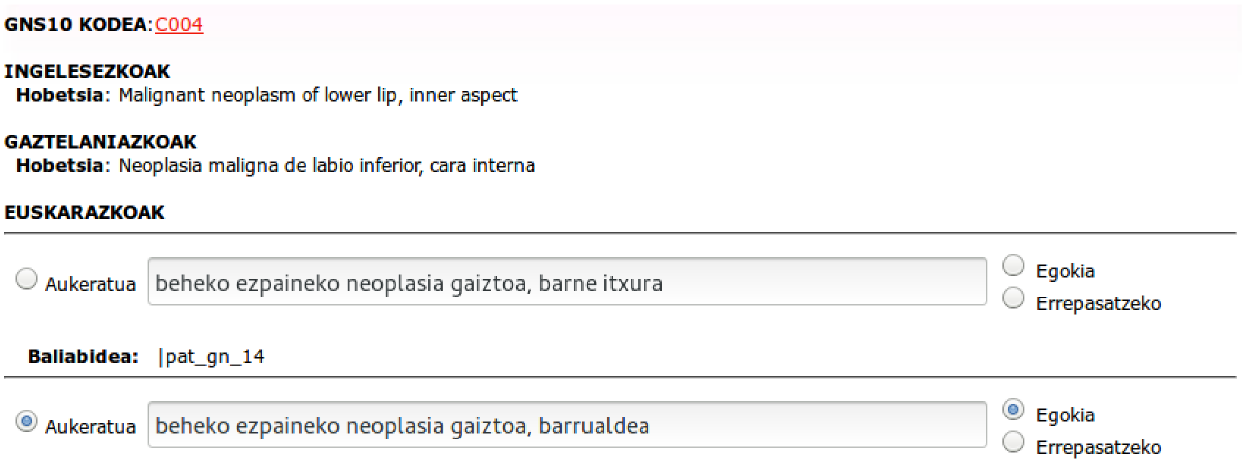

Figure 5: Screenshot of the web page for the validation of ICD-10-CM.

of the absolutive singular declension mark (row 2 in Table 2) when it is not needed.

- Edit distance 2: most of the cases were caused by an unnecessary plural mark in one of the words (row 3 in Table 2).

- Edit distance 3 and 4: in all the analysed cases it was made clear that the automatic translation has an incorrect mark of the genitive. Basque has two types of genitive that are equivalent to English preposition "of" . One is used to describe possession (possessive genitive, "-ren" mark) and the other location (locative genitive, "-ko" mark). By default, KabiTerm uses the possessive genitive to avoid massive overproduction as it is considered more general and in most of the cases it is the best option. The errors found with 3 and 4 edit distance are the descriptions in which the locative genitive should be used instead of the possessive one (row 4 in Table 2).

- Edit distance 5 and higher: From 5 edit distance on, the mistakes found are a combination of previously mentioned errors (row 5 in Table 2), or a bad choice of a nested term (row 6 in Table 2), or a combination of both.

Even if it remains impossible to measure the time saving for the generation of ICD-10-CM, the estimation is made considering the usual time spent on this kind of tasks and the time spent in this case, and the savings are around 4 times better than the time needed to translate the descriptions manually. For instance, the expert would need from 5 to 10 minutes to translate each code without any automatic translation, whereas with automatic translation the whole process takes from 1 to 2 minutes in case the translation is accurate.

4. Conclusions

In this paper, a system for the automatic generation of ICD-10-CM descriptions in a minor language such as Basque is

	Error type	Source	Correction
1	Spelling error	des**c**emetozelea, eskuineko begia	deszemetozelea, eskuineko begia
2	Singular article overproduction	atrofia optiko**a** primarioa, aldebikoa	atrofia optiko primarioa, aldebikoa
3	Incorrect plural mark	beste parafili**ak** batzuk	beste parafili batzuk
4	Incorrect genitive mark	kornea**ren** neoplasia gaizto	kornea**ko** neoplasia gaizto
5	Error combination	orbita**ren** zelulitis**ak**	orbita**ko** zelulitis
6	Incorrect nested term	**urradura**, eskuineko aldaka	**abrasioa**, eskuineko aldaka

Table 2: Examples of post-edited terms.

presented. The system called KabiTermICD is based on nested terms inside those descriptions and using translation patterns generates Basque equivalents.

As far as we know, this is the first work published on the automatic translation of ICD-10, and shows that automatic translation of ICD-10-CM is a promising investment to localise it. The work published here is specially useful for minor languages that can not afford a manual translation and its costs. Being a rule-based system, it that can be adapted to any other language pair, specially to the ones with a corpus not big enough to allow the implementation of corpus-based models. In any case, the lack of other systems makes the results not comparable at this point.

The results show the system to be robust enough, even if it is still open to improvements. Around 60% of correct translations proves a good accuracy in order to translate, and in the cases in which post-editing has been necessary, they were mostly small changes that did not require of big time spent by the expert.

For the future, in order to improve the precision of KabiTerm, the identified errors will be corrected, developing new patterns to improve its recall.

This work showed us that KabiTerm can be easily adapted to translate new medical vocabularies.

Bibliographical References

Abdoune, H., Merabti, T., Darmoni, S. J., and Joubert, M. (2011). Assisting the Translation of the CORE Subset of SNOMED CT Into French. In Anne Moen, et al., editors, *Studies in Health Technology and Informatics*, volume 169, pages 819–823.

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

Langlais, P., Yvon, F., and Zweigenbaum, P. (2008). Analogical translation of medical words in different languages. In *Advances in Natural Language Processing*, pages 284–295. Springer.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., and McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.

Perez-de Viñaspre, O. and Oronoz, M. (2014). Tanslating SNOMED CT Terminology into a Minor Language. In *Proceedings of the 5th International Workshop on Health Text Mining and Information Analysis (Louhi)*, pages 38–45, Gothenburg, Sweden, April. Association for Computational Linguistics.

Perez-de-Viñaspre, O., Oronoz, M., Agirrezabal, M., and Lersundi, M. (2013). A Finite-State Approach to Translate SNOMED CT Terms into Basque Using Medical Prefixes and Suffixes.

Perez-de Viñaspre, O. and Oronoz, M. (2015). SNOMED CT in a language isolate: an algorithm for a semiautomatic translation. *BMC medical informatics and decision making*, 15(2):S5.

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

Schulz, S., Bernhardt-Melischnig, J., Kreuzthaler, M., Daumke, P., and Boeker, M. (2013). Machine vs. Human Translation of SNOMED CT Terms. In C.U. Lehmann et al., editor, *MEDINFO 2013*, pages 581–584.

Zhu, Y., Pan, H., Zhou, L., Zhao, W., Chen, A., Andersen, U., Pan, S., Tian, L., and Lei, J. (2012). Translation and Localization of SNOMED CT in China: A Pilot Study. *Artificial Intelligence in Medicine*, 54(2):147–149, February.