EURALEX XIX

Congress of the European Association for Lexicography

Lexicography for inclusion

7-9 September 2021

Virtual

www.euralex2020.gr

Proceedings Book
Volume 2

Edited by Zoe Gavriilidou, Lydia Mitits, Spyros Kiosses

EURALEX Proceedings

ISSN 2521-7100

ISBN 978-618-85138-2-2

Published by: SynMorPhoSe Lab, Democritus University of Thrace

Komotini, Greece, 69100
e-edition

Publication is free of charge

Edited by: Zoe Gavriilidou, Lydia Mitits, Spyros Kiosses

English Language Proofreading: Lydia Mitits and Spyridon Kiosses

Technical Editor: Kyriakos Zagliveris

2021 Edition

A Lexicographic platform for migration terminology: problems and methods

Chiari I.[1]

Dipartimento di Lettere e Culture Moderne, Sapienza University, Rome, Italy
isabella.chiari@uniroma1.it

Abstract

"Language on the Fly" is a lexicographic resource for the domain of migration. The peculiarity of migration lexicon is due to scope (often geographical and institutional) and time. The language of migration is found on an international level where it is defined, for example, by institutions, like the EU regulations (both legal and administrative); on a national level, where general international procedures are modified and adapted to the specific country administrative and general migration policies and, finally, on an ordinary level which is interlinked to issues that migrants have to face in their interactions with institutions (social security, health, education, administrative issues). This paper focuses on corpus-based procedures used to build the second version made of a set of 2,094 entries and collocations starting from Italian language corpora specifically built to represent the three levels of lexicon, and further translated in 5 EU languages and 10 non-EU languages. The translation process also involves corpus-based techniques and multilingual corpora. Building the lemma list on three - specially built - Italian corpora using keyword extraction techniques, the glossary also uses corpus-based techniques to extract glosses that are further rewritten using controlled language in Italian in order to facilitate the use in cultural mediation contexts.

Keywords: migration lexicon; corpora; glossary; multilingual corpora; lexicography

1 Introduction

The "Language on the fly" project is an online platform which aims to provide reliable and updated linguistic, lexicographical, and documentary information for orientation language in the first reception of migrants, immigrants, and asylum seekers upon arrival in an EU country. The multimedia platform, accessible online also from mobile devices such as smartphones and tablets, and downloadable offline, will allow to frame the terminology useful for access, orientation, and reception of the country in three versions: a lexicographic reference resource; guide for humanitarian workers; guide for migrants. The prototype of the platform, to be released by the end 2021, covers basic local, national, and international terminology in the relating field in Italian, English, Arabic and French, while the general resource is built in order to cover many more target languages especially among the languages of countries sources of migration.

The project also addresses methodological questions about the development of linguistic-terminological resources related to the domain of migration, and investigation of linguistic and psycho-social needs for the orientation of the recipients of the platform. The peculiarity of migration lexicon is due to different aspects, mainly depending on the scope (often geographical and institutional) and time. More specifically the language of migration has an international or transnational level where it is defined for example by institutions, like the EU regulations (both legal and administrative); there is a further national level, where general international procedures are modified and adapted to the specific country administrative and general migration policies and a final ordinary level that is interlinked to issues that migrants have to face in their interactions with institutions (social security, health, education, administrative issues). All these aspects and more specifically the first two are furthermore constantly changing with frequent modifications in the legal framework on this subject (Dublin and its revisions, in Italy for example Security decrees that have changed asylum typologies in the last years, etc.). Furthermore, some of the terminology is also used in a non-technical sense in common media generating potentially dangerous ambiguity (*illegal and legal immigration, refugee, economic migrant*, etc.). This complexity needs to be addressed both for the updating of general lexicographic works, but also for the use of cultural mediators that are assisting migrants on their path in the new country and, finally, directly to migrants as beneficiaries of reference glossaries and guides.

The project stems from the need for inclusion that is a basic right especially for those who make significant sacrifices and face danger to reach a safe country and who are often penalized by the insufficient linguistic and legal information when facing the procedures of asylum and residence permit acquisition. This tool aims to fill the gap in reference tools in the hands of legal assistants and operators by providing a reliable resource, accessible and constantly updated and enriched with new languages.

The goal of the project is to bring together different types of operational, scientific, and psychosocial skills in a way to produce a model of linguistic guidance that is both synthetic and characterized by precision and translation accuracy, and usable for different types of publics (lexicographers, mediators and humanitarian workers, migrants). The constant interlocutors of the research group are in fact the civil society organizations operating in the humanitarian field in Italy and Europe, the operators, and the guests of the reception centres of Lazio, the Immigration Offices of the Provinces and the mediators that operate there, the voluntary associations operating in the assistance to the migrants. The prototype of the lexical resource is in fact based on the Italian case, providing a starting point for a quadrilingual resource (in its first version) with Italian, Arabic, English, and French languages.

The methodology used is particularly innovative as it combines essential aspects for the development of quality resources, complete and immediately usable for users. In particular, the first phase of experimentation related to the needs of the guests in reception centres through a questionnaire that has been proposed to guests of reception centres in the Rome area and the use of corpus and computational linguistic techniques for the selection of the terminology to be treated.
The first release of the "Language on the Fly" Italian glossary (multilingual in output) was based on three corpora representing the international, national, and ordinary levels of description of the migration lexicon, further elaborated to extract the relevant terminology with different statistical techniques and compared with previously available reference glossaries. The paper will illustrate the methodological structure and procedure to produce the lexicographic resource, its dynamic corpus-based methodology and some of the specific challenges that this kind of terminology poses to lexicographic description and its multiple uses.
The paper is organized in the following way: §2 provides a background on resources that address similar issues from different perspectives and that motivate the need of a new approach for the glossary that is proposed; §3 is a reflection on the different addressee of glossaries on the migration network and on ways to integrate focus on those differences in a unique resource; §4 illustrated the structure of the monolingual section of the glossary; §5 illustrated the cycles by which the SL lemma list is constantly enriched and checked; §6 describes the macro-structure of the construction of the TL section of the dictionary illustrating the asymmetrical relations in content.

2 The Background: Migration Lexicons and Phrasebooks

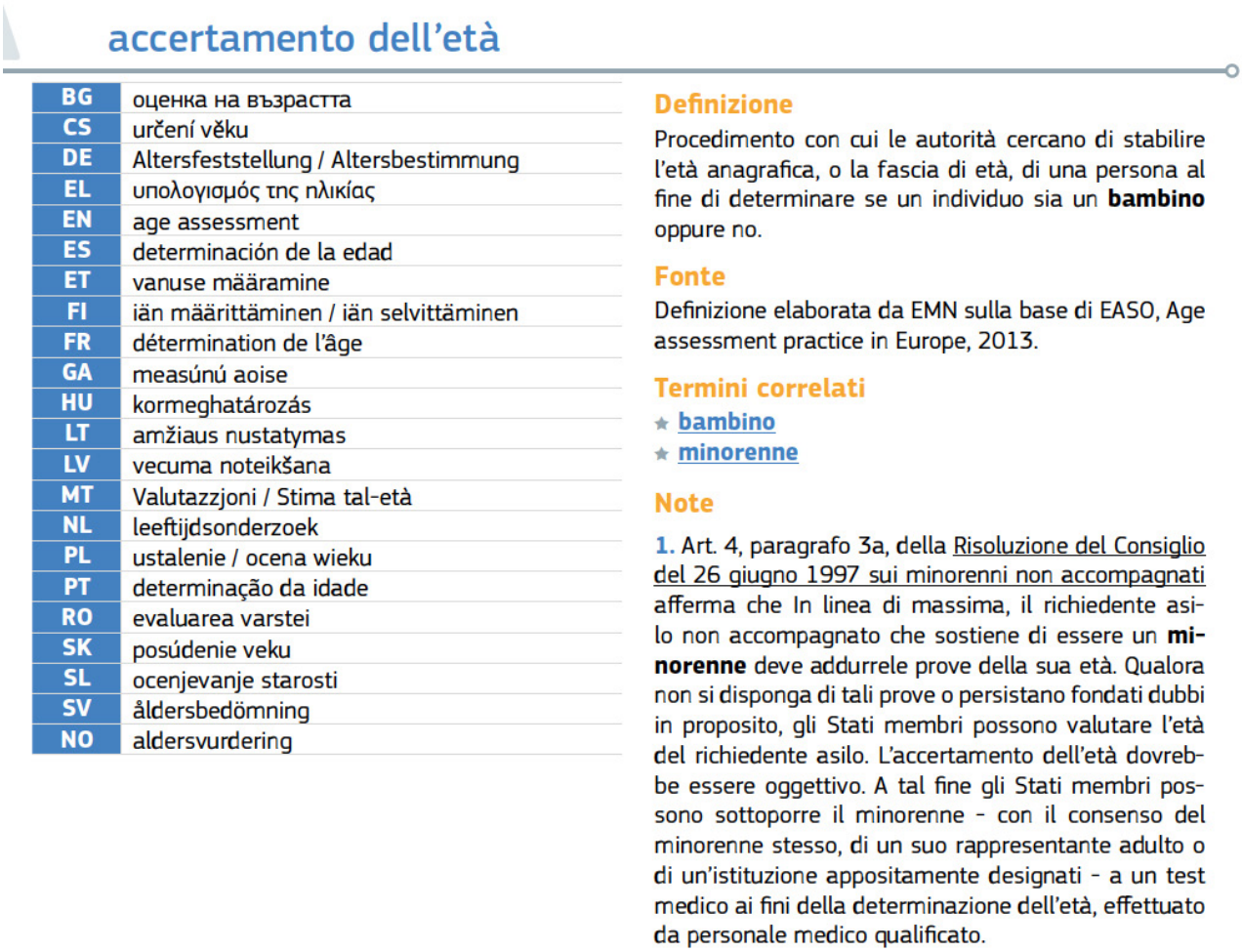
Lexicography strives to provide translations and cultural reference points, and the dissemination of 'emergency' resources such as online glossaries and phrasebooks, which are often -de facto- prepared by non-professionals without the necessary control of accuracy and in-depth analysis. Examples of this kind are *The Refugee Phrasebook* (Paul Feigelfeld 2016) that is a set of Google Doc Sheets including useful phrases (a general set of 600 phrases, 150 phrases for helpers) that are said to be covering 28 languages. The last update seems to date 2017. The approach is to cover general language basic needs as can be seen in Figure 1. There is also a specific small portion of 150 phrases dedicated to issues defined as generally *juridical* (Feigelfeld 2016).

SETTINGS LANGUAGES PRINT

ENGLISH	GERMAN [DEUTSCH]	ARABIC PHONETICS (FUSHA) / SYRIAN PHONETIC	FARSI [فارسی] PHONETIC
Hello	Hallo	Marhaba	Salaam
Welcome	Willkommen	ahlan wa sahlan	Khosh amadid
good morning	guten Morgen	Sabáh al-khayr	Sob bekheir
good evening	guten Abend	massa Alchayr Masa al-khayr	Shab bekheir
goodbye	auf Wiedersehen	ma'a 's-saláma / bye	khodaafez
sorry / excuse me	Entschuldigung	afwan, law samáht	Bebakhshid
please	bitte	lou tismah/Afwan	Lotfan
thank you / thanks	danke	Shukran	Merci
you're welcome [response to thank you / thanks]	gern geschehen	Ahlan wa sahlan	Khahesh Meekonam
my name is...	ich heiße...	ismi	Esmam ... ast
What is your name?	Wie heißen Sie?	Shou Esmak	esmetoon chieh?
I'm from...	Ich komme aus...	Ana min	Man az .... miyam
family	(die) Familie	Oussra / Aa'ila	Khaanevaadeh
this is my husband	das ist mein Mann	hada zawji / da zawji	Een shoharame

Figure 1: Example from *The Refugee Phrasebook.*

The same kind of approach that is mainly focused on Arabic for Syrian refugees is Gorsau (2015) and contains 200 Arabic sentences, questions and phrases translated into 26 European language, based on the idea of basic needs such as directions, food, accommodation, transport, but not covering any legal or administrative terminology regarding migration.
On the other hand, authoritative multilingual resources are mainly focused on legal and administrative aspects related to EU regulations or best practices such as those of the Council of Europe, and therefore not considering the deep changes that intervene in the 'systems' of reception of individual countries, or their specificities terminologies (European Migration Network 2018, International Organization for Migration (IOM) 2019, Perruchoud 2004), more specifically on Italian language (Programma Integra n/a, Anci, Cittalia & SPRAR 2015). Worth of a specific mention is the Glossary of the European Network *Asylum and Migration Glossary*, started in 2014 and with the last release in 2018 (6.0). It describes fully, with relevant referring sources, 256 entries in 23 languages of the EU. The aim of the glossary is institutional and is aimed at members of the EU and policy makers.



accertamento dell'età

BG	оценка на възрастта
CS	určení věku
DE	Altersfeststellung / Altersbestimmung
EL	υπολογισμός της ηλικίας
EN	age assessment
ES	determinación de la edad
ET	vanuse määramine
FI	iän määrittäminen / iän selvittäminen
FR	détermination de l'âge
GA	measúnú aoise
HU	kormeghatározás
LT	amžiaus nustatymas
LV	vecuma noteikšana
MT	Valutazzjoni / Stima tal-età
NL	leeftijdsonderzoek
PL	ustalenie / ocena wieku
PT	determinação da idade
RO	evaluarea varstei
SK	posúdenie veku
SL	ocenjevanje starosti
SV	åldersbedömning
NO	aldersvurdering

Definizione

Procedimento con cui le autorità cercano di stabilire l'età anagrafica, o la fascia di età, di una persona al fine di determinare se un individuo sia un **bambino** oppure no.

Fonte

Definizione elaborata da EMN sulla base di EASO, Age assessment practice in Europe, 2013.

Termini correlati

- bambino
- minorenne

Note

1. Art. 4, paragrafo 3a, della Risoluzione del Consiglio del 26 giugno 1997 sui minorenni non accompagnati afferma che In linea di massima, il richiedente asilo non accompagnato che sostiene di essere un **minorenne** deve addurrele prove della sua età. Qualora non si disponga di tali prove o persistano fondati dubbi in proposito, gli Stati membri possono valutare l'età del richiedente asilo. L'accertamento dell'età dovrebbe essere oggettivo. A tal fine gli Stati membri possono sottoporre il minorenne - con il consenso del minorenne stesso, di un suo rappresentante adulto o di un'istituzione appositamente designati - a un test medico ai fini della determinazione dell'età, effettuato da personale medico qualificato.

Figure 2: Example of entry *accertamento dell'età* in EMN Glossary (Italian version, 6.0, 2018).

Most of the existing resources are generally centered on EU languages and do not contain useful elements for interfacing with the languages of the beneficiaries of the reception and this provides a strong limit to the use of the tools themselves in the real application field by translators, humanitarian mediators, and operators. A similar approach but with a broader scope covering general EU terminology is that of (EUROVOC 2010, IATE 2018, UNHCR 2006), containing, as a section, sets of entries pertaining to the domain of migration.
Of different character are works dedicated to the words describing immigration as associated to racism and xenophobia, also relevant and interesting for the topic but slightly out of focus for the purpose of this research (Bolaffi, Gindro & Tentori 1998, Bhopal 2004). Also, a different focus is that of Małgorzata (2016) that bears educational aims trying to describe visually concepts related to migration, forced migration, torture and related issues.

3 A Resource for Whom?

As mentioned above, and shown by the diversity of approaches and addressees of previous works, the challenges posed by the migration lexicon are based on its internal stratification depending on factors such as: national, regional, and local differences, diversity in possible audiences (from institutional international stakeholders to aid workers and finally migrants themselves) and from the intrinsic non correspondence of different migration responses along with rapid changes in legislation for each recipient country taken into consideration.
At a macro level we can organize the migration lexicon into three large sub-sets or domains differentiated by scope:

a) An international or transnational level, here identified with the institution of the European Union's regulations (legal and administrative/institutional and its migration approach principles, e.g., Dublin regulations and its revisions). This domain is generally best described since it needs to be standardized at least for all EU languages (although not taking into consideration the languages of migrants which are seldom corresponding to any of the above-mentioned languages) – a selection of terms of this typology is that provided by the (European Migration Network 2018) that describe terms such as *Accordo di Cotonou* ('Cotonou Agreement), *adozione fittizia* ('adoption of convenience'), *lavoratore migrante* ('migrant worker'), *minore non accompagnato* ('non accompanied minor'), *paese di transito* ('country of transit'), *permesso di soggiorno* ('residence permit'), etc. that are horizontal to the documentation found in specific EU National regulations;

b) The second macro sub-set is national in scope, and it regards procedures, regulations, adaptations, and additions that are modified and proposed by each specific country administrative and general migration policies. Each country does in fact implement and define regulations and implementations in individual and not cross-nationally comparable ways. These regulations and implementing decrees are constantly changing depending on government direction and overturns and on public opinion stances (e.g., in Italy for example Security decrees that have deeply changed in time asylum typologies in the last few years, etc.). The second layer is country-specific and often does not offer any official or non-official translation of its content and terminology being of national, regional, or local interest. In the case of Italian, examples are *soccorso civile* ('civil aid'), *maggiorenne* ('adult'), *lettera di assunzione* ('hiring contract'), *profugo* ('refugee'), *a tempo determinato* ('fixed term'), *Decreto Sicurezza* ('Security Decree'), etc.;
c) The last sub-set of the lexicon relevant to migration management process concerns activities that are interlinked to aspects that migrants must face in their interactions with institutions (social security, health, education, administrative issues).

For all these sub-strata (which in actual use can also overlap, especially between level A and B) there is an urgent need to develop strategies and tools that cover significant gaps both at international and national levels.

The basic needs regarding glossary enrichment concern: methodology (corpus-based, corpus design and updating); inclusion of languages of migrants; inclusion of the overall stratification of domains pertaining migration not focusing only on institutional and regulative texts. Furthermore, the migration domain is affected by a widespread media usage of terms in a non-technical sense generating potentially dangerous ambiguity (*illegal and legal immigration, refugee, economic migrant*, etc.). This complexity needs to be addressed both for the updating of general lexicographic works, but also for the use of cultural mediators that are assisting migrants in their path in the new country and finally directly to migrants as beneficiaries of reference glossaries and guides.

Thus, a general resource on migration terminology serves multiple beneficiaries: general lexicographers, policy makers, aid workers and mediators, migrants themselves. This aim can be achieved by devising a source language description that is accurate but understandable, with explicit references and a focus that includes as target languages the languages of migration specifically connected to a single reception country. Furthermore, the second macro-level (national or country-based) requires the glossary to be monodirectional and poses great challenges for translation equivalents since terminology is often not internationally or transnationally shared. The "Language on the Fly" project is by design monodirectional, since it aims to consider especially the lexicon that is used in country specific national, regional, and local texts on migration and that often do not conform to international standards. The prototype itself can nevertheless be applied to any source country (and language) as a methodology and procedure for collection and description of data.

The concept behind the project and its novelty is both in trying to address a very complex stratification of lexicon that is not standardized and that, by definition includes languages that are not only typologically different from most EU languages but that are also characterized by administrative and legislative as well as social differences that are challenged in translation of official documentation. Furthermore, in some cases terminology is completely absent since most TL are not countries that do not have a significant incoming tradition in immigration so can lack sufficient corpus materials to provide comparable terminology and need compromises, paraphrases and glosses in order to be properly used. So, lexicological analysis and translation require consistent, explicit approaches and a full and comprehensive knowledge of the underlying context of the countries that represent the TL (a point particularly critical in the case of Arabic).

From a macro-structural point of view the main innovations in concept, compared to the available resources are:

a) Diverse language inclusion with a focus of TL of countries of origin of migration – not only EU languages;
b) Taking into account the time variable both at national and international level that often requires a temporal specification for definitions, since legislation on the subject has constantly changed; this also require accessible strategies to represent those changes;
c) A wider audience that varies from EU policy makers on the one hand to migrants themselves on the other, requiring complex choices in accessibility and levels of representation, and that are motivated both by linguistic and civil purposes of the work.

This prototype suggests some significant theoretical points regarding multilingual lexicography. One of the main issues regards language directions and the relationship between languages and countries where the language is spoken, since in the case of domains other than general language of more standardized portion of the lexicon, the relationship with terminology to the overall country management, and legislative and administrative apparatus is very tight. So, a relevant element of reflection regards the nature of linguistic resources themselves linked to areas similar to that of migration, that has a crucial international role, needs multilingual resources, but resources cannot be conceived as traditional *linguistic* and *translation* issues that can be represented bi-directionally without posing significant threats to the domains represented.

This theoretical point is not specific to the domain of migration, but emerges in a striking way in this domain since migration requires tools and linguistic resources that cover languages and cultures that come into contact and conflict in terms of responsibility and humanity of the addressee of documents and cultural mediation, adding to a known asymmetry the burden of bearing strong consequences on the lives of people accessing texts and asking to have their rights recognized and granted.

4 Structure of the Monolingual Section (SL)

The prototype model proposed, to be released by the end 2021, covers basic local, national, and international terminology

in the relating field in Italian (as source language) and English and Arabic as target languages. Working versions are now being developed in six EU languages (Italian, English, French, Spanish, Portuguese, German) and 10 non-EU languages (Arabic, Azerbaijani, Serbian, Pashto, Russian, Persian, Albanian, Turkish, Chinese, Norwegian). Priority is given to languages that are most represented as countries of origin of migration in Italy.
The LoF project is organized as a processing cycle starting from the source language (SL), which in the prototype is Italian, and the further linking to target languages (TL) linguistic data, which in the first release will be Arabic and English. The phases of processing are organized in cycles to benefit from the empirical approach and to take into account updates in relevant documents and procedures to be considered.
For the monolingual description of lemmas (whose selection process is described in the next paragraph) the scheme of the prototype is exemplified in Figure 2.

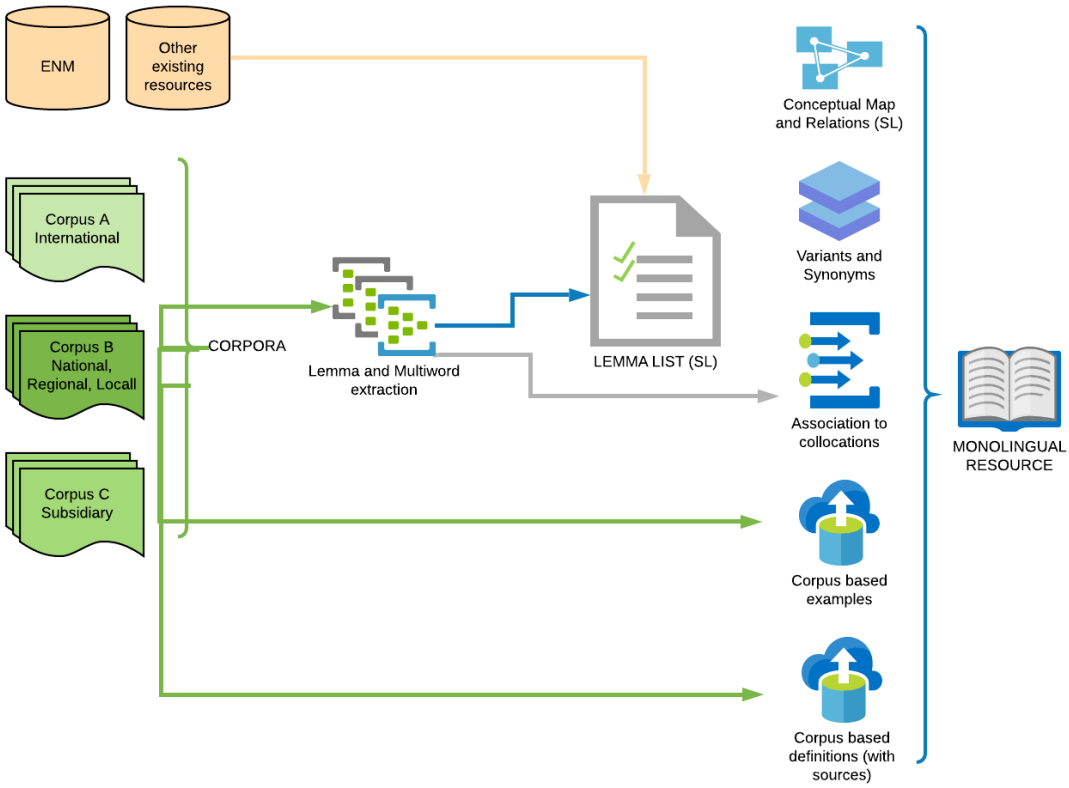


Figure 3: Monolingual SL Resource Map.

5 Building the Lemma List

As can be observed in Figure 3, the first step in glossary building is the definition of a starting lemma list in the SL. The lemma list is further increased by progressive cycles deriving from semi-automatic terminology extraction by the three Italian corpora, which are also continuously updated. As shown, the starting point is a set of lemmas already collected in previous general glossaries including Italian as SL or TL (564 lemmas).
The core of the new lemma list extraction is corpus based. Three corpora representing the three levels of stratifications have been constructed and are constantly updated. An overview of current characteristics from the corpora is:

Corpus Title	Area	Occurrences	Text Typologies
A.1 Corpus of International EU Migration (Italian)	EU	ca 5,000,000	Documentation from the European Parliament, The Council of Europe, The Court of Justice, The European Commission on migration issues
B.1 Corpus of National Migration 1 (Italian)	Italy	ca 5,000,000	Italian legislation and documentation of migration management specific to the Italian context
C.1 Corpus of Subsidiary National Migration 1 (Italian)	Italy	ca 5,000,000	Documentation for social security, health, education, administrative issues etc.

Table 1: Lof Italian SL Corpora.

Lemmas and multiwords (including collocations) are extracted from each corpus, checked by concordance, and inserted into the LoF lemma list (which is defined by a version number depending on cycles of corpus monitor). The release of the

LoF resource is (provisionally) composed of 2,094 entries (of which 1,558 multiwords, and 89 named entities).

Following the definition of the entry list a set of procedures are applied for the description of the SL entry properties:

a) Identification of relationships (semantic and conceptual) among lemmas;
b) Identification of formal variants and synonyms;
c) Identification and association of collocations and idioms;
d) Extraction and selection of usage examples from corpora;
e) Extraction of primary source definition for key concepts with relative source.

Steps b) c) d) and e) are all performed by accessing build corpora for SL. The result of these operations is converted in a relational database form that produces a monolingual resource of terminology that will be the starting point for the processing for all target languages. For pure exemplification the entry for the SL has the following structure (but multiple layouts depending on the device and purpose of the applications), see Figure 4.

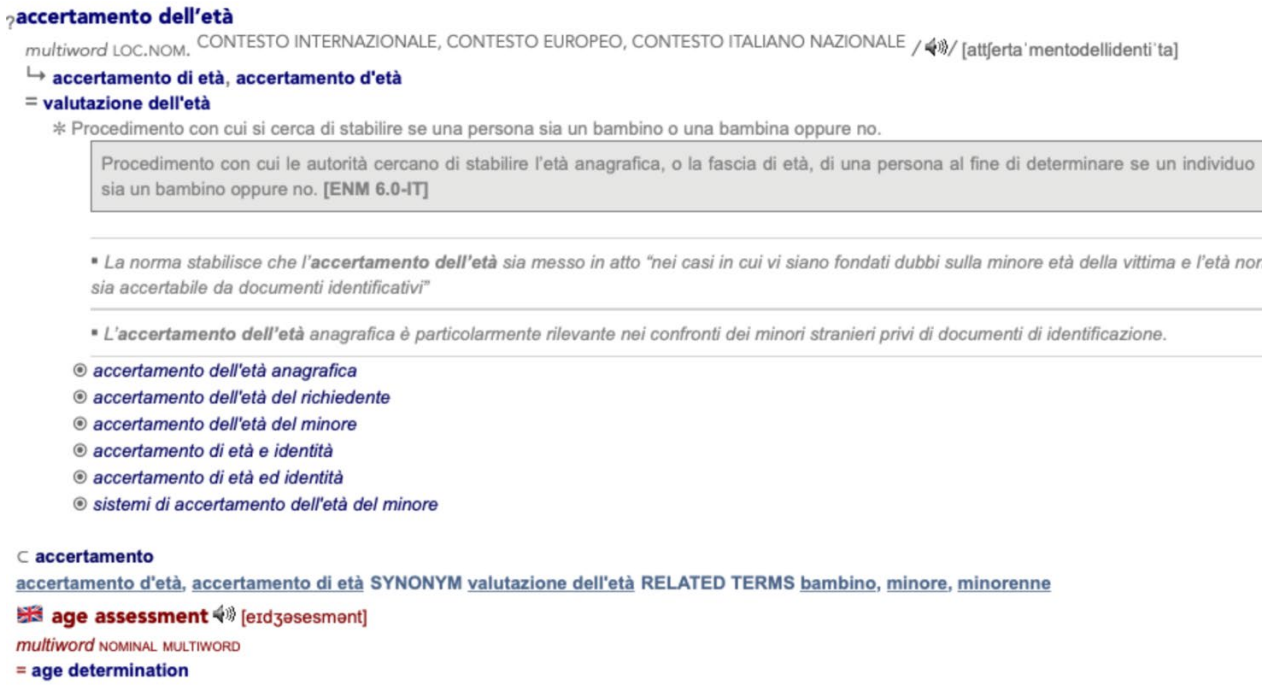


Figure 4: Example *accertamento dell'età* in Monolingual SL resource database.

The provisional layout provides grammatical classes, information about the general scope (international, national, common language), pronunciation both as playable sound files and in phonetic transcription. The structure of the glossary further provides formal variants (such as for *accertamento dell'età, accertamento di età, accertamento d'età).* Formal variants have been introduced not only to take into account variability but also to enable the resource to be used potentially in text mining and in computational tools. Where available, synonyms have been provided (as in this case *valutazione dell'età*). The definition preceded by the ❊ symbol is provided in simple controlled language to be fully understandable by average users, using a prototypical approach, while the technical definition – where available - is provided in a box along with reference to the original source. Examples are all corpus based and are not manipulated in any way but chosen in order to represent common context usages of the entry, see example 1.

(1) L'accertamento dell'età anagrafica è particolarmente rilevante nei confronti dei minori stranieri privi di documenti di identificazione. [*age assessment is particularly relevant for foreign minors without identification documents*]

Finally, a list of collocations found in the corpus is given along with semantic relations and related terms, as can also be observed in Figure 5.

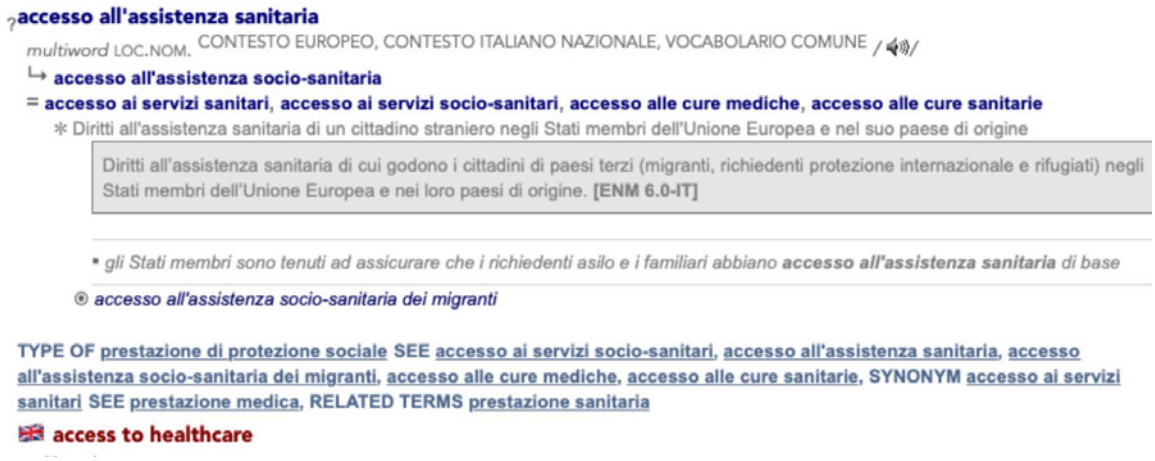


Figure 5: Example *accesso all'assistenza sanitaria* in Monolingual SL resource database.

6 Building the TL Macro-Structure

The following steps regarding the work on TL involves different procedures depending on languages and availability of corpus resources to rely on for the translations, mapping, examples and definitions. For Corpus A (International EU) we have built parallel corpora at least for the EU languages included in the project and comparable corpora for non-EU languages, where possible, facing a number of challenges. The sub-set of the lexicon extracted from Corpus A is generally more widespread and standardized, so the process of extracting TL entries, synonyms, examples and definitions from parallel or comparable corpora poses less of a challenge. To this set belong entries such as: *Accordo di Cotonou, beneficiario di protezione internazionale, cittadino di un paese terzo, Convenzione di Dublino, discriminazione diretta, migrante di seconda generazione*, etc.

One of the main challenges remains the absence of a common migration framework that grant good quality translations and do not introduce potentially risky ambiguities. In this respect glosses are provided for translations in cases where the cultural and legislative background demands it.

Processing of Corpus B (National, Regional and Local Scope) is the biggest challenge since it contains terminology, concepts and entities that are country specific and that derange significantly even from the EU given framework. This peculiarity pushes to rely on new terms in TL in order not to generate ambiguity with close or similar terms common to the international terminology. Examples of entries extracted from the B corpus of Italian are: *Agenzia del demanio, ente territoriale, certificazione sanitaria, ufficio G.I.P., SPRAR, abuso della libera circolazione, associazioni del Terzo Settore, autonomie locali*, etc.

Corpus C (Subsidiary) concerns additional aspects in migration management which are not directly linked to the process of asylum and are related to issues like social security, health, education, and administrative issues. From the point of view of corpus collection and comparative or parallel corpora build it can be considered of medium complexity since in many countries forms and infos about these subject matters are also available in different language and are generally less sensible in content and nature (e.g., *carta di identità, stato civile, certificate di matrimonio, tessera sanitaria*, etc.).

At the moment of writing the quantitative data about entries description is the following:

Languages (ISO codes)	# entries	of which # multiwords	variants	synonyms
IT	2,094	1,558	455	970
EN	2,094	1,609	519	1,196
AR	1,009	764	0	22

Table 2: Lof entries described for SL and TLs.

The structure of TL entries linked to the SL entries is the same as the monolingual side. This uncovers many differences in word usages and helps identify asymmetric relations among entries, both simple lemmas and multiwords. For example, observing the English collocations associated with *age assessment* we find many that are not represented in the Italian corpus and in the entry such as *errors in age assessment; means of age assessment; to make age assessment; age assessment procedures; to carry out the age assessment; age assessment guidance; the process of age assessment; integrated age assessment*. We also find different synonyms to be taken into account such as *age determination*. A similar situation happens with the correspondent Arabic where *accertamento dell'età* can be translated by two multiwords تقدير السنّ or تقدير العمر. In this case there is no substantial difference in the two and the terminology is fairly standardized in Arabic. A slightly different case is that of *permesso di soggiorno* that in English is the standard *residence permit* while in Arabic, although strongly associated to تصريح الإقامة in a general sense, can also be found in corpora a form less standardized as إذن الإقامة that would be "permission to reside" or "authorization of residence" along with *residence*

permit.

7 Challenges

The process of building a corpus based (country-based) glossary of migration terminology has proven to provide many challenges in the linking of data and especially in working on languages that possess a completely different administrative, legislative, and generally cultural background. The case of Arabic is particularly significant since it includes dozens of linguistic varieties being spoken by around 422 million speakers in 25 different countries. The challenge is not only cultural but specifically linguistic since some expressions are peculiar only to some areas and not to other, although are still considered Standard Arabic (and not dialectal forms). The linking between the monolingual SL structure to the TLs structures is still to be perfected to assure the user the capability of moving themselves through the different information provided in the entry and their translation equivalents.

The most significant base for new entries extraction has been Corpus B, since it is the corpus that is strongly related to the migration process description but nevertheless contains terminology, which is not shared in EU or international documentation, since it depends on the country choices in legislation and administration and at the same time very rarely possesses official translations in any language other than the SL (both EU and non-EU languages).

Working on this kind of terminology further exposes significant problems in the difference in policies that EU countries adopt as reflected by the language chosen and by the political implication of administrative and legislative determinations. Thus, working on migration terminology has proven to be a challenge that needs to be further taken into account to provide better services and to assure the democratic access to basic human rights, which are also guaranteed by linguistic choices and accessible description.

8 References

Bhopal, R. (2004). Glossary of terms relating to ethnicity and race: for reflection and debate. In *Journal of Epidemiology & Community Health*, 58(6), pp. 441-445.

Bolaffi, G., Gindro, S. and Tentori, T. (1998). *Dizionario della diversità: le parole dell'immigrazione, del razzismo e della xenofobia*. Firenze: Liberal Libri.

European Migration Network (2018). *Asylum and Migration. Glossary 6.0*. Belgium: European Migration Network (EMN).

EuroVoc Thesaurus - The EU's Multilingual Thesaurus. Accessed at: https://op.europa.eu/en/web/eu-vocabularies/th-dataset/-/resource/dataset/eurovoc [18/04/2021].

Feigelfeld, P., (2016) *Refugee Phrasebook - Juridicial Phrases for Refugees*. Accessed at: https://www.refugeephrasebook.de/ [18/04/2021].

Glossario. Accessed at: http://www.programmaintegra.it/wp/risorse/glossario/ [18/04/2021].

Gorsau, H. (2015). *Special dictionary for Syrian refugees, migrants and asylum seekers travelling towards Europe*. St-Orens, France: Editions Goursau.

Interactive Terminology for Europe. Accessed at: https://iate.europa.eu/home [18/04/2021].

International Organization for Migration (IOM) (2019). *International Migration Law N°34 - Glossary on Migration*. Interactive Terminology for Europe: International Organization for Migration (IOM).

Małgorzata, T. (2016). *Visual Icon Dictionary on Migration*. Poland: One World Association

Perruchoud, R. (2004). *International migration law: Glossary on migration*. International Organization for Migration.

The Refugee Phrasebook. Accessed at: https://en.wikibooks.org/wiki/Refugee_Phrasebook [25/03/2021].

UNHCR (2006). *UNHCR Master of Glossary of Terms*. Geneva 2, Switzerland: U. N. H. C. f. Refugees.