Online Dictionaries for immigrants in Greece: Overcoming the Communication barriers

Anna Vacalopoulou, Voula Giouli, Maria Giagkou and Eleni Efthimiou
Institute for Language and Speech Processing, R.C. "Athena"
Artemidos 6 & Epidavrou, Maroussi, Greece
Email: {avacalop; voula; mgiagkou; eleni_e}@ilsp-athena-innovation.gr

Abstract

In this paper we describe on-going work aimed at the creation of a suite of specialized Language Resources (LRs) intended for users not previously targeted at, namely, adult immigrants in Greece. The ultimate goal being to help them integrate in the Greek society, we aim to provide support touching at basic linguistic, social and everyday issues. The suite comprises: (a) bilingual dictionaries integrating a grammar component; (b) sample typical dialogues, relevant to communicative situations that the target group is most likely to cope with; and (c) a multilingual parallel text corpus that adheres to domains that are of interest to the target group. These LRs will be integrated into a web interface coupled with advanced search mechanisms that will provide innovative accessibility options for visually impaired users. The paper describes the intended LRs suite elaborating on the corpus compilation and processing, as well as on the dictionaries macro- and micro-structure the focus being on the methodological principles underlying selection and organization of the dictionary entries.

Keywords: bilingual dictionaries; immigrants in Greece; user needs and requirements; corpora

1. Introduction

The unprecedented growth of immigrant population in Greece over the last decade has led to the adoption of policies aimed at their smooth integration and social inclusion. In this context, language education plays a central role within the action plans and measures taken. The document reports on work still in progress within the framework of *eMiLang* project. It elaborates on the bilingual dictionaries that are being developed in order to support the communicative needs of immigrants in Greece. Section 2 outlines the project scope and aims, which ultimately guided *dictionary design* and *platform functionalities*, whereas Section 3 briefly describes the target group with respect to main characteristics, and their needs and requirements. Section 4 elaborates on the dictionary specifications (macrostructure and microstructure), the emphasis being on the provisions taken towards addressing the specificities of the target group. The lexicographic considerations taken into account in lemma selection are presented in section 5, whereas the primary data (corpora) that comprise the sets of bilingual textual collection and the methodology adopted for collecting them are presented in section 6. In the last section, we present conclusions and future work.

2. The framework: eMiLang project

The *eMiLang* project aims to develop a *digital infrastructure* tailored to support adult immigrants in Greece to overcome the communication barriers in their everyday interactions, and in administrative, social and educational settings. The ultimate goal is to assist both immigrants and policy makers in their joint efforts for smooth integration of the target groups to the Greek society. The intended infrastructure encompasses two inter-related pillars: (a) the development of LRs, namely **specialized multilingual parallel corpora** in the form of informative material and **bilingual dictionaries** (*partly* extracted from these corpora), and (b) the implementation of a multilingual, multimedia web **interface** designed so as to integrate the digital content (dictionaries and informative material). This interface will also offer advanced search mechanisms and information retrieval capabilities. Finally, a news aggregator will be integrated into the system, offering digital information services to the users.

3. The target group: needs and requirements

As it is evident, dictionary design in terms of language coverage, entry selection and presentation mode is *user-oriented*. The user perspective in dictionary making is considered along the following axes: (a) users' *reference needs*; (b) their *proficiency level* and *background knowledge*; (c) their reference *skills* and *strategies*; and (d) effectiveness of dictionary use training (Varantola, 2002).

To infer the needs and requirements of the target group, an investigation had to be conducted in order to primarily identify their profile(s) and respective needs. One major difficulty in this task was the inability to perform proper analyses employing appropriately designed questionnaires and tests as proposed by mainstream lexicographic research (Atkins, 1998). This was mainly due to the fact that locating the intended users at such an early stage of the dictionary-making process and persuading them to participate in any type of survey was extremely difficult, since these people whose upmost concern is to struggle for a living in a new and unknown environment. Instead, we opted for postponing any immediate contact with the target group until a first version of the platform was available for on-line user feedback elicitation during the pilot use phase. In fact,

this approach is consistent with what has been called "simultaneous feedback" from the target users to the compilers (De Schryver et al., 2000). Based on the assumption that in most cases, user feedback usually comes too late because it can at best be considered for implementation in the revised edition of the dictionary, this approach caters for the identification of prospect user needs and preferences by launching/testing pre-final dictionaries coupled with questionnaires. In this way, hypotheses may be tested, and refinements and modifications can be implemented where needed and in the light of feedback obtained by the users during dictionary development.

Thus, to initialize dictionary compilation, consultation of official, general-purpose statistical data took place. As a matter of fact, there are relatively few data available detailing immigrants in Greece and their characteristics. And apart from the sparse quantitative and qualitative surveys on immigration (Baldwin-Edwards, 2004; 2008), the only sources available were the *2001 Census* survey, along with figures obtained from *Eurostat (*http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat). Information thus obtained reveals the principal immigrant nationalities in Greece as being *Albanian*, *Bulgarian*, *Georgian*, *Romanian*, *Russian*, *Ukrainian*, *Polish*, *Pakistani*, and *Egyptian*. Moreover, regarding age, the vast majority of immigrants fall within the range 15-64. Dependent employment has also been recorded as the principal reason for award of residence permits (68% of the total). Following this, roughly equal at 12% each, are family reunification and self-employment, whereas very few immigrants enter Greece for study purposes. Moreover, as far as *immigrants' presence in the Greek labour market* is concerned, Census data regarding male immigrants' main occupations, the principal employment has been in building construction, followed by agriculture, industry and tourism. Female employment is dominated by occupations such as housekeeping and cleaning, and also employment in sectors such as agriculture, and tourism.

The characteristics of the immigrant population with respect to educational level and language literacy were also obtained from the aforementioned sources. According to *Census 2001* survey and *Eurostat* data, the vast majority of immigrants in Greece are of educational levels ranging from medium to law. More specifically, statistical data show that immigrants in Greece mainly fall in one of the following three groups with this respect. The first group comprises immigrants who have completed secondary education before entering the country; the second class consists of those who failed to progress beyond primary school. Both classes are populated with people originating from European countries (Albania, Bulgaria, Poland, and Serbia). The last one, which comprises immigrants from countries in Africa and Asia, includes those who are classed as illiterate. Additionally, most of the immigrants were reported as having little or no knowledge of the Greek language prior to entering the country. Moreover, one can safely infer that the target group has almost no prior experience in the use of dictionaries or other linguistic resources, and that they have low to medium level of computer literacy.

Another interesting fact was the highly visible increase in the number of immigrant children recorded in state schools, especially since the mid-1990s. And although this seems to be a difficult problem to tackle per se, the difficulties immigrant parents face when communicating with their children's tutors have been reported as also being a problematic issue (Baldwin-Edwards, 2008).

From all the above, we conclude that the intended target group is diverse in terms of nationality, level of literacy and language proficiency in Greek, yet the tendency is for lower a level. On the basis of users' profile, their needs and requirements were identified or inferred. The ultimate goal being to help immigrants integrate in the Greek society, we aim to provide support touching at basic linguistic, social and everyday issues along the following axes:

(a) communicative needs in official settings (as for example, in dealing with the Greek authorities, applying for a green card, etc.);
(b) communicative needs in social settings;
(c) communicative needs in order to cope with every-day issues (as for example travelling and transportation, etc.);
(d) language learning in formal or informal settings;
(e) familiarization with the general cultural and social context.

4. Dictionary Specifications

The nationalities identified determined the languages to be covered by the bilingual dictionaries, namely: Greek–Albanian (EL-AL), Greek–Arabic (EL-AR), Greek–Bulgarian (EL-BG), Greek–Chinese (EL-CH), Greek–English (EL-EN), Greek–Polish (EL-PL), Greek–Romanian (EL-RO), Greek–Russian (EL-RU), and Greek–Serbian (EL-SR).

Furthermore, the specifications recommended that the whole process of dictionary compilation be corpus-based; this refers to headword selection (in order to identify the appropriate vocabulary), sense selection and distinction, and collocations and usage examples extraction. Finally, the specifications stressed the importance of user-friendliness of the dictionary and ease of access as a basic feature of the underlying platform, since it addresses the needs of people as outlined above and, to this end, meta-language should be kept to a minimum.

In the following sections the implementation of these basic guidelines will be presented in more detail.

4.1 Macrostructure

4.1.1 Types of entries

Each bilingual dictionary will comprise around 15,000 entries which cover mainly the basic vocabulary of Greek. And although a formal complete list of basic vocabulary is still missing for the Greek language, in the current implementation, the basic vocabulary is conceived as one which comprises not only the most frequent items but also less frequent words and phrases that are relative to everyday life.

Another substantial category of lemmas is the one often occurring in official, administrative or other documents that the target group is likely to come up with during their stay in Greece, as for example when applying for a residence permit, etc. To this end, selected technical vocabulary, that is, terms pertaining to domains/subject fields that are of utmost interest to the target group have been included as well.

Because of the fact that the target group is generally expected to lack basic encyclopaedic information about Greece, this dictionary also contains proper nouns. These include the names of: (a) geographical entities (i.e., cities, islands, regions etc.), (b) official bodies (i.e., ministries and other official organisations), and (c) geopolitical entities (*Ηνωμένα Έθνη = United Nations*). Both official bodies and geopolitical entities are quite often expressed by acronyms which are also retained in the lemma list.

In terms of form, this dictionary contains two main categories of entries: single-word and multi-word lemmas. Multi-word entries may include expressions (*καλό ταξίδι = have a nice trip*), collocations (*χαρτί υγείας = toilet paper*), etc.

Different alternatives[1] of the same word or phrase are separate entries which are interlinked with each other. For instance, *Ολυμπιακοί Αγώνες* (Olympic Games) and *Ολυμπιακοί* (Olympics) are two separate dictionary entries linking to each other. Similarly, *κινητό τηλέφωνο* (mobile phone) and *κινητό* (mobile) are also listed as stated above. The most 'complete' form of such entries is considered as the main entry and contains the rest of the information in this dictionary. The secondary entry/entries serve as cross-references to the main entry. When two entries linked by cross-reference belong to different registers, the main entry is the most formal type, as more likely to occur in official and/or state documents. In the case of acronyms, the main entry is the full name of the entity (*Ευρωπαϊκή Ένωση = European Union*), the acronym (*ΕΕ = EU*) being a cross-reference. Acronyms are normalised for easy reference and are thus all written without dots among letters.

[1] For alternative spellings of words or phrases, the main entry follows the official school grammar spelling whereas other spellings are cross-references.

Although none of the cross-references are fully developed entries, certain types of information are included, so that users can access them immediately, without having to follow the cross-reference link. These are: hyphenation, pronunciation, link between masculine and feminine nouns, the three forms of adjectives (i.e. masculine, feminine and neutral) and domain (see section 4.2).

4.1.2 Lemma distinction

The main criterion for lemma distinction is morphology. Thus, the following are listed as different entries: *Οκτώβριος* and *Οκτώβρης (=October), μέρα* and *ημέρα (=day), εβδομάδα* and *βδομάδα (=week)*. The next criterion for lemma distinction is part of speech. Thus, homographs belonging to different parts of speech form separate entries (*άρρωστος, άρρωστη, άρρωστο = ill*, *άρρωστος = patient*). Because of the difficulties arising from the fact that Greek is a highly inflectional language, the past participle of verbs is treated lexicographically as an adjective, thus forming a separate entry (*αγαπημένος*, *αγαπημένη*, *αγαπημένο = beloved or favourite, p.p. of the verb αγαπώ = love*; *χαμένος*, *χαμένη*, *χαμένο = lost, p.p. of the verb χάνω = lose*).

Along similar lines, all types of word derivatives are separate entries. Thus, adverbs (*αργά = slowly, διαφορετικά = differently*) are different entries from the respective adjectives (*αργός, αργή, αργό = slow, διαφορετικός, διαφορετική, διαφορετικό = different*).
All single-word entries appear in the 'base' form, in the way that would be expected to appear in regular monolingual dictionaries: for verbs, this is the first person singular present in the active voice; for nouns, the singular nominative; for adjectives and past participles, the nominative positive (in this case, in the masculine, feminine and neutral); for adverbs, the positive. Exceptions to the above rules occur when what is usually considered as the 'base' form is either ungrammatical or particularly infrequent in Greek (*χιονίζει = it snows*, the third instead of the first person, *λεφτά = money*, the plural instead of the singular, *συναχώνομαι = catch a cold* instead of *συναχώνω = cause somebody to catch a cold*).

Nouns referring to professions or other people's activities form two different entries (masculine and feminine) as, in most cases, their morphology differs (*δάσκαλος* and *δασκάλα = teacher, ιδιοκτήτης* and *ιδιοκτήτρια = owner, ταξιτζής* and *ταξιτζού = taxi driver*). Exceptions to the above rule would be nouns with identical masculine and feminine forms (*ηθοποιός = actor* and *actress*; *ταμίας = male* or *female cashier*; *υπουργός= male* or *female minister*).

The comparative and superlative of certain highly frequent adjectives and adverbs also form separate entries. Thus, *λιγότερος*, *λιγότερη*, *λιγότερο = less* as well as *περισσότερος*, *περισσότερη*, *περισσότερο = more* appear separately from *λίγος = little* and *πολύς = much*, respectively.

4.2 Microstructure

4.2.1 Meanings and examples of usage

As this dictionary is mainly targeted toward starter learners of Greek who are in need of speedy learning, it has been decided that only basic meanings would be included in it. Meanings however are neither defined nor directly translated; they are implicitly presented through one or more examples of usage, which bear the informative load. Examples of usage are thus a core element of the dictionary.

Furthermore, examples in this dictionary are carefully selected so as to reflect not only the different meanings but also the most basic forms of usage, grammar and/or collocation. Thus, for instance, the active and passive of verbs are presented in separates when voice differentiates meaning as well; the same stands for verbs used with different prepositions etc.

As the emphasis of this dictionary has been to include as much information as possible but in the most user-friendly way possible, examples have been selected so as to be as interesting as possible to the target group. To this end, a combination of different corpora has been used. A large part of the examples for the basic vocabulary was extracted from the Hellenic National Corpus, although usually shortened and/or simplified to suit the target group level.

In terms of length, examples are short and contain no excess information. They usually consist of one simple sentence, although some dialogue is included to exemplify everyday phrases, such as greetings or asking for information. Apart from accelerating the learning process, the brevity criterion also simplifies the ambitious work of translating everything into 9 languages.

As it is customary in most multilingual dictionaries, examples also play the role of describing each meaning, due to lack of definition. This has placed additional difficulty in selecting the right example for each meaning. For instance, an example of the verb *αγωνίζομαι = struggle* would be *Αγωνίστηκε πολύ, για να καταφέρει αυτό που ήθελε = She struggled a lot to get what she wanted*.

Last but not least, taking into account the great variety of backgrounds from which the target group of this dictionary comes, extra care has been taken toward political correctness. All examples are free of any social, political, racial, national, and religious or gender bias.

4.2.2 Communicative/subject domains

Each meaning/example of the entry words is categorized in broad domains that reflect certain communicative contexts an immigrant in Greece may be involved in. As noted above immigrants are a special case of language learners, i.e. their needs are those of a summer week tourist and of an active citizen at the same time. An immigrant has, for example, to go shopping or book an apartment, to register a child in a public school and object to the employer when labour legislation is violated. In this view, the domains have to be detailed enough to cover as many possible different communicative needs and comprehensive enough to facilitate usability. An additional factor that has led to the categorisation of entries into domains is that, according to studies, users rarely go through the list of senses for each dictionary entry, usually selecting the first meaning (Lew, 2004). It is, therefore, suggested that users will be more likely to identify the appropriate meaning of multi-sense entries when these are clearly categorised into domains. These domains are:

- **Education**, e.g. *πανεπιστήμιο* (university), *ΑΕΙ* (acronym for Higher Education Institution)
- **Labour – insurance**, e.g. *επίδομα ανεργίας* (unemployment allowance), *ημερομίσθιο* (wage)
- **Law, justice and public safety**, e.g. *ποινικός κώδικας* (penal code), *δικηγόρος* (lawyer)
- **Finance**, i.e. anything related to money and the economy, including taxation, bank transactions etc.
- **Public administration – politics**, i.e. vocabulary that does not fall into any of the above categories and concerns administration, bureaucracy, the government, the political framework etc., for example, *βουλή* (parliament), *πιστοποιητικό οικογενειακής κατάστασης* (civil status certificate)
- **Transportation and travel**, i.e. vocabulary related to urban transport and travelling in general
- **Geography**, which will include an extensive list of countries, nationalities and languages, as well as all the major Greek cities and areas
- **Physical condition and health**, i.e. parts of the body, diseases, doctors, etc.
- **Science and technology**, i.e. computers and technological gadgets, some widely used scientific fields and terms
- **Environment**, i.e. flora and fauna, geomorphology, weather, ecology, etc.
- **Culture, recreation and the media**, i.e. vocabulary from the arts, hobbies and spare time, television and the media in general
- **Relations – family**, i.e. words for family and social relations
- **House and accommodation**, i.e. parts of a house, furniture and appliance, as well as vocabulary relevant to accommodation in general, e.g. hotels and rooms to let.
- **Public holidays and Greek traditions**, comprising the most common Greek holidays and celebrations, as well as culture specific traditions that an immigrant is unfamiliar with.

Finally, the most populated domain is, as expected, **general vocabulary**. For educational reasons mainly, part of the general vocabulary will be further

subcategorized into distinctive vocabulary groups such as:

- numbers
- clothing and accessories
- food and cooking
- time
- space
- colours
- measurement units
- everyday interaction (informal words and expressions).

4.2.3 Additional entry information

Information for each dictionary entry includes phonetic transcription, pronunciation (audio file), hyphenation, alternative entry types (cross-references), elementary grammatical information (i.e. the masculine, feminine and neutral type for all adjectives and past participles) and examples of usage. Each example is translated into 9 languages, with the entry word/phrase highlighted in the example.

Apart from entries themselves, examples are also pronounced in Greek and Bulgarian using a synthetic voice. This is meant to help people with vision or literacy problems on the one hand and the vast majority of people who are not familiar with the Greek alphabet on the other.

As far as hyphenation is concerned, it is included for all single-word entries, in an attempt to help users compose hand-written or electronic texts. What is more, hyphenation in Greek is not arbitrary, so this feature is expected to help more advanced users familiarise themselves with the basic rules of hyphenation in Greek.
All multi-word entries are interlinked with each of their components (excluding functional words such as prepositions, conjunctions and articles). Not only does this feature help in easy reference, but it also has a pedagogical added value, as most of the words contained in phrases are inflected types of lemmas. Thus, users are guided to link each individual type to the base form of the lemma.

5. Lemma Selection Methodology

Dictionary entries were semi-automatically selected from a variety of sources, including (a) a large (POS-tagged and lemmatized) reference corpus of the Greek language, namely the Hellenic National Corpus (http://hnc.ilsp.gr/), (b) the Greek counterpart of the specialized multilingual parallel corpus, and (c) from already existing dictionaries and glossaries, customized to better suit the user needs (communicative situations and relevant vocabulary, etc.).
As it has been noted above, a proportion of the entries is part of what can be conceived as the *basic vocabulary* of Greek. This does not only mean the most frequent items attested in the HNC, but also less frequent words and phrases that are relative to everyday life, and which are used to populate the domains described above (such as *μαξιλαροθήκη = pillowcase* or *πάνα = nappy*).

Similarly, a corpus-based methodology has been employed for the semi-automatic selection of entries which belong to a more technical vocabulary with the use of NLP lingware (see section 7 below), coupled with manual correction and selection of the most frequent/appropriate terms.

Finally, this dictionary follows the closed vocabulary concept, thus including every word in the examples as an entry itself for easy reference. This has led to adding a considerable amount of entries ad hoc and keeping a better balance, in terms of content, between everyday vocabulary and the administrative jargon of the public service.

6. Corpora and Linguistic Processing

The role of corpora in the project is two-fold: (a) to provide linguistic evidence and aid linguistic introspection and (b) to form the multi-lingual informative textual material. Two types of corpora were consulted in this respect: the Hellenic National Corpus (HNC), a large reference corpus of the Greek language, and the *eMiLang* specialized corpus. As it has already been mentioned, the former was used to extract the source language (EL) material (headword selection, sense discrimination, usage examples), whereas the latter has already been used for the extraction of terms adhering to the domains catered for in the project. Additionally, it will form the data pool for the development of the informative multilingual material.

More precisely, the *eMiLang corpus* comprises texts that adhere to domains that are of interest to the target group, namely: *administrative/legal*, *health, education, transport* and *civilization*. The texts have been selected from various sources over the Internet: official websites of public bodies, organizations, the EU portal, etc. This corpus currently amounts to 172K words. As far as balance is concerned, this was achieved for the domains *health, education* and *transport* (c. 30K words each). The domain *administrative/legal* outperformed the other three (c. 95K words) because of the availability of data and data sources. Data pertaining to the *civilization* domain were the most difficult to collect (only 17K words) due to the strict Intellectual Property Restrictions, and they were only kept for the off-line part of the corpus.

One peculiarity of the textual collection at hand is that it will also form the data pool for the creation of the informative material. To this end, documents containing information that is dated or obsolete, were only retained to form the off-line corpus from which linguistic evidence was extracted, and they were appropriately marked so as to be used with consciousness thereof.
A metadata scheme for the efficient representation of the corpus data along with the encoding of the linguistic annotations has been implemented for the efficient

management and retrieval of the textual data. This scheme is compliant to widely accepted standards so as to ensure reusability of the resource at hand, namely the specifications of the Text Encoding Initiative (TEI). Metadata elements have been deployed which encode information necessary for text indexing with respect to text title, source, author, publication date, etc. (bibliographical information) and for the classification of each text according to text type/genre and topic. Metadata elements for catalogue descriptions compatible with the specifications proposed by the ISLE[2] Meta Data Initiative (IMDI) were also added manually to the whole corpus in view of rendering the corpus searchable by prospect users.

After text selection and documentation, extended manual validation (where appropriate) was performed. Normalization of the primary data was kept to a minimum so as to cater, for example, for the anonymisation of the official documents (that is the deletion of person names) and for the conversion of collected files to a format appropriate for further processing.

Text processing was then applied via an existing pipeline of shallow processing tools for the Greek language (Papageorgiou et al., 2002). These processing steps include: (a) part-of-speech (POS) tagging and lemmatization; (b) Named Entity (NE) Recognition; and (c) Term extraction. The Greek POS-tagger has been developed in-house and is based on Transformation Based Learning architecture. Trained on Greek textual data from various sources (newspapers, internet, etc.) it assigns Part-of-speech labels to words in a sentence. Following POS tagging, lemmas retrieved from a Greek morphological lexicon were assigned to every word form. At the next stage, Named Entity Recognition was performed on a subpart of the corpus using a Maximum Entropy Named Entity Recognizer (MENER), a system compatible with the Automatic Content Extraction (ACE) scheme (http://www.itl.nist.gov/iad/mig/tests/ace/), catering for the recognition and classification of the following types of NEs: person (PER), organization (ORG), location (LOC) and geopolitical entity (GPE). For the purposes of the current project only NEs of the types (LOC) and (ORG) were retained. A Greek Term Extractor (TE) was finally used for spotting terms and idiomatic words. TE proceeds in three pipelined stages: (a) morphosyntactic annotation of the domain/specialised corpus, (b) corpus parsing, i.e., identification of syntactic constituents using a pattern grammar endowed with regular expressions and feature-structure unification, and (c) lemmatization. The tool employs a hybrid methodology, in that statistical evaluation of candidate terms skims valid domain terms, lessening, thus, the over-generation effect caused by pattern grammars.

[2] International Standards for Language Engineering

7. Conclusions and Future Work

We have hereby presented work still in progress targeted at the development of on-line dictionaries for immigrants in Greece. Initial considerations involve entry selection and entry organisation the ultimate goal being to better suit the intended users' needs.

Future work involves the implementation of a platform that will be user-friendly, featuring search functionalities for easy access to the entry via the lemma or the word form. To this end, a tool will be integrated, which links each inflected form to a very large morphological lexicon of Greek. This is expected to be of enormous help to the lookup process. Moreover, fuzzy-matching techniques will also be employed, and users who misspell words will be presented with a list of correct spelling alternatives from which they can choose. This is one of the features adding to the pedagogical nature of this dictionary.

8. Acknowledgements

The research within the project "*eMiLang: Improved Informative Digital Services Targeted on Immigrants*" leading to these results has received funding from the Greek state, under the framework of *O. P. Digital Convergence (2007-2013)*.

9. References

Atkins, B.T.S. (1998). *Using Dictionaries: Studies of Dictionary use by Language Learners and Translators*. Tübingen: Max Niemeyer Verlag.

Baldwin-Edwards, M. (2008). *Immigrants in Greece: Characteristics and Issues of regional distribution*. MMO Working Paper No. 10, Jan. 2008.

Baldwin-Edwards, M. (2004). *Statistical Data on Immigrants in Greece*. Athens: Mediterranean Migration Observatory and IMEPO.

De Schryver, G.M., Prinsloo, D.J. (2000). Dictionary-Making Process with 'Simultaneous Feedback' from the Target Users to the Compilers. *Lexikos 10*, pp. 1–31.

IMDI, Metadata Elements for Catalogue Descriptions, Version 2.1, June 2001.

Lew, R. (2004). *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-Bilingual Dictionaries by Polish Learners of English*. Poznań: Motivex.

Papageorgiou, H., Prokopidis, P., Giouli, V., Demiros, I., Konstantinidis, A. & Piperidis, S. (2002). Multi-level, XML-based Corpus Annotation. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, Las Palmas, Spain, pp. 1723-1728.

TEI Guidelines for Electronic Text Encoding and Interchange (http://www.tei-c.org).

Varantola, K. (2002). Use and Usability of Dictionaries: Common Sense and Context Sensibility?. In M.-H. Correard (ed.) *Lexicography and Natural Language Processing: A Festschrift in Honour of B. T. S. Atkins*. Grenoble, France: EURALEX, 2002.