

A MODERN MULTILINGUAL MULTIMEDIA DICTIONARY FOR CHILDREN

GEORGIOS KARAYANNIS & MARIANNA KATSOYANNOU

Speech Processing Institute

1. INTRODUCTION

The topic of this paper is the presentation of the first dictionary of the *Lexipaedia* series. It refers to a general language dictionary that is directed towards elementary school students, aged 10 to 12, for whom Greek is their mother tongue. In parallel, the dictionary's design foresees foreign language coverage, given that multilingualism is not only a necessity in our age, but a civil and cultural choice on the European continent. The Speech Processing Institute has a tradition in the production of technological products for children and *Lexipaedia* is the first systematic attempt in the field of multilingual multimedia dictionaries. The laboratory prototype presented here has been developed to address the requirements of the full series of dictionaries, directed towards users of differing age groups; in addition to *Lexipaedia for Elementary School* the plan foresees *Lexipaedia for Junior High School* and *Lexipaedia for High School*.

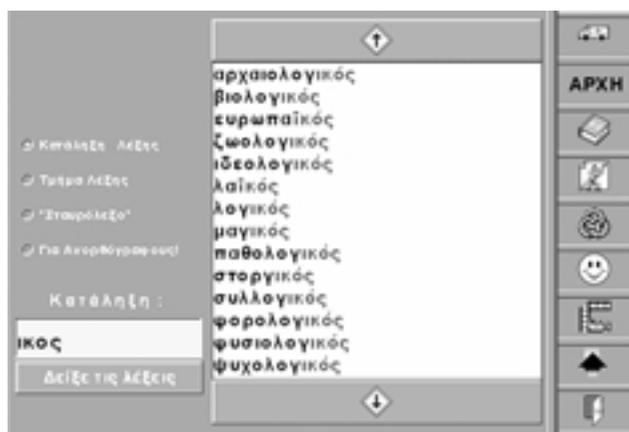
In *Lexipaedia for Elementary School*, special importance is paid to ergonomics, so as to make it attractive to children, and simplify the entire working environment. The macrostructure consists of approximately 8,000 lemmata; the programme includes different educational games that test language knowledge and abilities, designed around the idea of a sea adventure. The formulations of the definitions are rigorous and precise, though the examples are chosen, wherever possible, from daily childhood routines, including schoolbooks. The comments possess a predominantly phrasal character, although they also include simple etymological comments. The junior high school level dictionary, with 10,000 lemmata, is being designed with an emphasis on collation of definitions in similar groupings and in topics, and with the externalization of definitive fields that are utilized during the processing of lemmata. Thus, there is an attempt to increase the users' interest in terminology. The term definitions have a more rigorous formulation than that of the elementary school dictionary, so that they can also be used for the study of the underlying subject matter. Finally, the high school dictionary, which contains 15,000 lemmata, has an almost academic character. Here, advanced tech-

nological techniques will be applied with simulated experiments in the physical sciences that will facilitate searching, by providing differences in meaning between related words.

Given that a dictionary is a reference point for the proper usage of language, its introduction in the educational process is important. It is well known though, that the Greek student body is not especially familiar with the use of printed dictionaries, although on the other hand, in our age, alternative learning methods, dominated by images, have been developed. We know that multimedia products can be especially engaging, that they can motivate the user and can facilitate learning, making it more experience-focused. This is the aim of *Lexipaedia*, through which there is an attempt to create both a work and a game environment simultaneously, in this way maintaining the interest of the child for a longer time span and being useful as a means of feedback in the acquisition of words, as well as a supplementary learning tool.

Beyond all this, however, the programme design has adopted a series of technological innovations, which makes it even more interesting. The manner in which a user is given information on the dictionary lemmata is based on the utilization of multimedia technology: *Lexipaedia* includes images, motion, sounds, and video, which are used both in the presentation of the meanings, as well as in the different games that accompany the language material. For each single word definition, a translation is provided in the following languages: English, French, German, Spanish, Russian, and Bulgarian. The use of a morphological dictionary allows for the provision of not only the complete paradigm of all the entries, but also all the basic morphological information for those words that are not treated in the dictionary. Special software allows the automatic phonetic transcription of words to be shown, while a second programme is utilized for hyphenation purposes. The use of an artificial voice is also important since it allows the vocalization of all lemmata, those included in the list of lemmata, as well as the remaining words that are supported only by the morphological dictionary. Most of the innovations referred to have already been included in the laboratory prototype, although some exist as subsystems, which are to be later included.

Following is a description of the programme usage by the student, and subsequently the lexicographic design itself. The various search methods for a lemma in the *Lexipaedia* deserve particular mention. To begin with, there is the usual method: the user types in a word, the word appears on the right hand side of the screen (adjacent words appear in alphabetical order in the same window) and also by clicking a screen-card appears that contains the meaning and the rest of the information. There are alternative search methods: the user can type in a suffix and view all the words with that particular ending, or another part of the word. For example, the following images show the search for suffixes and the results (reverse dictionary).

1. Search suffix *-ώνω*2. Search suffix *-ικός*

An interesting search method, temporarily named “Crossword”, allows for the identification of certain digraphs: the user types in the word he is interested in, placing a question mark in the place of the letter that he is unsure about. In response, the system displays the form (or forms) of the word that has been identified on the screen: for example, the user that hesitates between *αδελφός* and *αδερφός* or *στεναχώρια* and *στενοχώρια* will receive two answers. The same applies for words such as *βρόμικος* and *βρώμικος* or *χλομός* and *χλωμός*, which continue to have two correct spellings. The two answers lead in each case to the same screen-card, which contains both spellings of the word accompanied by a

common definition. This search method is more interesting with homonyms: for example if the user does not recall how the word /'klisi/ is written, he enters: κ λ ? σ η and is advised of the existence of one entry κλήση and two entries κλίση, with differing meanings. Here a clarification is in place: the above search method applies to words with equal amount of letters; whenever there is a difference pertaining to a diphthong or grouping of consonants, the word search is not possible. This problem is solved with an alternative search method that can locate the entry regardless of its spelling: After the user has chosen the word of interest using one of the aforementioned search methods, he passes to the screen-card that contains information on the definition of the following form:



In what follows, we shall deal with the organization of the information contained on such a screen.

2. LEXICOGRAPHICAL PLANNING

2.1 SELECTION AND CATEGORIZATION OF ENTRIES

The initial entry catalogue that is included in the dictionary was compiled using schoolbook glossaries. The macrostructure was created in the following manner:

1. It was supplemented by using certain sources that we have at our disposal which pertain to the frequency of a word's occurrence and the basic vocabulary. The sources are the following:
 - Frequency statistics produced by the Speech Processing Institute,
 - The Basic Dictionary of Oral Language, from the research of A. Hatjisava (doctoral dissertation),
 - The prescribed elementary school dictionary, from the word catalogues compiled at the Pedagogical Institute.

2. It was modified in accordance with certain selection criteria, some traditionally lexicographic and others related to the electronic form of the dictionary. For example, word forms produced by derivation (such as the noun diminutives or the comparative degree of adjectives) are not listed separately, since this information is given automatically through a specific function of the dictionary. The same applies to participles ending in *-μένος* since a function that provides the full inflection of the lemma is provided. In addition, unidiomatic or dialectical forms are avoided, without however, ignoring words which appear in known proverbs and stereotypical phrases.

The aforementioned procedures resulted in a basic list of lemmata, which were reviewed and supplemented using the following methods:

1. By making new lemmata out of the definitions, to ensure that all words utilized will constitute lemmata in the dictionary. New lemmata (which are usually absent from frequency catalogues) are added by this method and are necessary for the already existing definitions. To give some examples, the word *άρθρωση* was added to the body of lemmata as a hyperonym of *αγκώνας* and *γόνατο*; the word *μύκητας* was added to define *μανιτάρι*, and so on.
2. Classifying the vocabulary material according to semantic fields. The primary purpose of this classification was to allow the editors of the dictionary to avoid processing the lemmata in alphabetical order, so as to avoid repetitions and circularity in definitions of terms. It was also used though, as a control device, so as to supplement the (possible) semantic omissions. For example, in the Table, the list of lemmata is given for all words related to the semantic field “*αίσθηση*” (‘sense’) where the words that were added to the initial body of entries by means of this procedural control are noted in italics. This process has also helped to supplement the entire range of the parts of speech, so that all the grammatical categories that are derived from the same root are represented in the dictionary. Hence, the noun *ζήλεια* (‘jealousy’) was added since the adjective *ζηλιάρης* (‘jealous’) was already included in the body of entries, *βραβεύω* (‘to reward’), since *βραβείο* (‘reward’) was included, and so on.

Excluding purely lexical lemmata, for which classification is effected in accordance with the example in the Table (p. 226), special categories of lemmata are created which in themselves constitute particular groups:

- Multiword units of vocabulary such as *ουράνιο τόξο* (‘rainbow’) or *παιδική χαρά* (‘playground’), which are included in the dictionary as autonomous entries, but at the same time, are connected with the words of which they are made up.

TABLE

Noun	Verb	Adjective	Noun
αίσθηση	αντιλαμβάνομαι αισθάνομαι-νιώθω	αισθητός	αίσθημα
		ευαίσθητος αναίσθητος ανθεκτικός	ευαισθησία αναισθησία
	λιποθυμώ		
	πονώ-πονάω		
ακοή αφτί=αυτί	ακούω κουφαίνω	κουφός	
αφή δέρμα πετσί	αγγίζω – εγγίζω ακουμπάω-ώ πιάνω πιέζω τσούζω ζεσταίνω κρυώνω πουντιάζω παγώνω τουρτουρίζω	απαλός πιεστικός καυτός ζεστός χλιαρός κρύος ψυχρός	κρύο ζέστη
γεύση γλώσσα	γλείφω	νόστιμος γευστικός άγευστος άνοστος ξινός αλμυρός=αρμυρός πικρός γλυκός γλυφός στυφός	νοστιμιά
όραση μάτι- οφθαλμός	βλέπω κοιτάζω – κοιτάω-ώ τυφλώνω αντικρίζω αγναντεύω	φανερός αφανής άφαντος ορατός αόρατος αόμματος τυφλός στραβός αντικρινός	βλέμμα ματιά
όσφρηση μύτη ρουθούνι	οσμίζομαι – οσφραίνομαι μυρίζω μοσχομυρίζω μοσχοβολάω-ώ μοσκοβολάω-ώ βρωμάω ρουθουνίζω -άω-ώ		μυρωδιά οσμή

emotion

Human body

The arrows ▼ and ► signify the transition to a different semantic field

- The entries that belong to closed grammatical categories or which form series. These include prepositions, conjunctions, pronouns, certain adjectives such as *μόνος, ίδιος, δικός*, numerals, lemmata for which the definition requires deictic elements (*εγώ, εδώ...*), also proper names (until now the dictionary has included names such as countries, continents, oceans, days of the week, months, and holidays) and in general all types of lemmata which could be included in general dictionary texts are included here.

2.2 STRUCTURE AND ORGANIZATION OF THE DEFINITION

The information provided for each lemma constitutes equivalent fields of a Data Base and is organized in the following manner:

2.2.1 *Spelling, Pronunciation and Morphological Information*

For each word-entry, all spelling variations are provided especially when common, e.g. *αυγό* and *αβγό* ('egg'), *αυτί* and *αφτί* ('ear'). As stated above, regardless of the form that the user entered during the search, the system will refer to a screen-card where the word is listed with both spelling forms. The spelling is accompanied by the hyphenation of both written forms. Phonetic transcription is given, despite the fact that all entries are voiced, enabling the user to listen to the word pronounced by simply pressing a key. The purpose of the transcription is to familiarize the child with the phonetic alphabet, which he can utilize during the learning of a foreign language.

Information on the grammatical category follows, in which the gender of nouns, or the conjugation of verbs is given. The problem of adjectives which are used as nouns is usually solved by creating a separate entry: The noun *ναυτικός* ('sailor') constitutes a different entry from the adjective *ναυτικός, -ή, -ό* ('naval'), *θαλασσινό* ('seafood'), is separate from *θαλασσινός, -ή, -ό* ('of the sea'), etc. There is no reference to inflectional tables since the user can observe the full word paradigm by pressing the correct key.

2.2.2 *Definition and Classification of Meanings*

Of course, the greatest weight is given to the formulation of the definition. In order for the definition to satisfy the needs of the user (comprehension or recognition of the meaning of a word), the word to be defined and its proposed semantic equivalent should not share a common etymology, so that we do not end up with tautological definitions. When no other solution exists –this occurs mainly in definitions of adjectives and adverbs– we provide for the creation of automatic reference to the corresponding entry. The same solution is applied in the case of certain nouns of action and whose definition includes the corresponding verb,

for example the word *ψάρεμα* ('fishing'), which is defined as *το να ψαρεύεις* ('to fish'), with automatic reference to *ψαρεύω* ('I fish').

The definition is expressed in complete sentences, avoiding mere adducing of synonyms; normally the definition consists of the hyperonym of the entry accompanied by its distinctive characteristics. The characteristics which, though important for the comprehension of a word, are not mandatory, are described as "συνήθη" ('usual') for example *ζάχαρη* ('sugar') = "...συνήθως σε μορφή σκόνης" ('usually in powdered form...'), though that which deviates from the logic of the definition can appear in the example (e.g. the characterization of a dog as "πιστός" ('faithful'), of a fox as "πονηρή" ('sly'). In addition, an attempt is made so that the word which is given as a semantic equivalent of the entry belongs to the same grammatical category (noun = noun, verb = verb). The main exception to the above rule pertains to adjectives, for which the definition usually consists of a relative clause that is introduced with "που". Another characteristic of the definition, mentioned here because it is not found in all children's dictionaries, is that it consists of a single sentence.

The definition can include indications of the cognitive area to which a word or one of its meanings belongs: Grammar, Mathematics, Physics, etc. Until now, we have been accustomed to encountering these characterizations in parentheses; in *Lexipaedia* they will be represented by icons.

The order of meanings is not determined on the basis of historical evolution. Usually the definition of the most frequent or the most specific or of the literal meaning precedes. One problem that remains is how to order verb meanings: given that the different syntactical usage of a verb usually produces differing definitions, the data entry operator is forced to choose one of two different methods of ordering: the first being the aforementioned, the second taking into account the number of definitions (where the intransitive use of the verb is first stated).

2.2.3 *Synonyms, Antonyms and Comments*

For each lemma meaning, words of similar meanings are provided under the heading "synonyms", as are the antonyms. The user can refer automatically to each one of these words and view its definition.

The card may also record comments, which include mainly stylistic and phrasal indications or grammatical comments, for example "does not form a plural" or "used primarily in literature". Here certain simple observations of an etymological nature are also included such as: "*ο δείκτης* ('index finger') λέγεται έτσι γιατί τον χρησιμοποιούμε για να δείχνουμε ('to indicate')" or "*η λέξη πλήκτρο* ('key') προέρχεται από το ρήμα *πλήττω* ('to pound') που σημαίνει *χτυπώ* ('to strike')".

2.2.4 Usage Examples, Expressions and Proverbs

The examples –which are usually taken either from schoolbooks or the everyday life of students– can contain additional information, which has no place in the definition and/or adds characteristics useful for the comprehension of a meaning, e.g. the example that corresponds to the lemma *ερπετό* (‘reptile’) adds “πολλαπλασιάζονται με αυγά” (‘propagated by eggs’). Particular importance is given to the phraseology framework in which a lemma is usually found and which is not always included in the definition, e.g. for a word such as *ακτινοβολία* (‘radiation’), an attempt is made for the verb *εκπέμπω* (‘transmit’) to appear in the example. The examples always consist of complete sentences.

In a separate Data Base field, proverbs and stereotypical phrases are noted, the innovation being that they are connected with all the words that are referred to. Thus, the expression “λύνει και δένει” appears in both of the verbs *λύνω* (‘untie’) and *δένω* (‘tie’), the expression “η ισχύς εν τη ενώσει” is given in the entries *ισχύς* (‘strength’) and *ένωση* (‘union’), etc. These expressions are always accompanied by their meanings, and, when it is deemed necessary, by comments on their usage.

Translation YANNIS GETSOS

References

- BEJOINT, H. & P. THOIRON. 1996. *Les dictionnaires bilingues*. Paris: Editions Duculot.
- Εκπαιδευτικό Πολυλεξικό: Προδιαγραφές απαιτήσεων-ανάλυση αναγκών. 1995. Athens: Douka Schools & Speech Processing Institute.
- I FONTANALS, J.R., ed. 1996. *Diccionari de Frequències*, vol. I. Barcelona: Institut d’ Estudis Catalans.
- Λεξικό της Νέας Ελληνικής Γλώσσας (δείγμα: ψηφία ζ, η, θ, ι). 1987. Thessaloniki: Institute of Modern Greek Studies-Manolis Triandafyllidis Foundation.
- PICOCHÉ, J. 1996. *Précis de lexicographie française; l’ étude et l’ enseignement du vocabulaire*. Paris: Nathan Université.
- SINCLAIR, J., ed. 1988². *Collins Cobuild, English Language Dictionary*. London: Collins.
- VOSTANTZOGLOU, T. 1990². *Αντιλεξικόν ή ονομαστικόν της νεοελληνικής γλώσσας*. Athens.

Optical Disks

- Dorling Kindersley Multimedia. *The Dorling Kindersley Children’s Dictionary*.
Speech Processing Institute. *Λογομάθεια. Διδασκαλία της σύγχρονης ελληνικής γλώσσας*.
Microsoft Explorapedia. *Children’s Interactive Encyclopedia*.