Performance Evaluation of Italian Electronic Dictionaries: User's Needs and Requirements

Isabella Chiari
Dipartimento di Studi Filologici, Linguistici e Letterari (DSFLL),
Università di Roma "La Sapienza", P.le Aldo Moro, 5, III Piano
Edificio ex Facoltà di Lettere e Filosofia, 00185 Roma

Abstract
Electronic dictionaries are nowadays to be considered as distinct and partially autonomous tools for common users, teachers and researchers. Information coded in the printed edition is extraordinarily increased and strengthened by querying functions implemented in the electronic version. Six of the major Italian human readable electronic dictionaries (ED) have been analyzed in order to observe capabilities of data extraction from the printed version, to estimate performance capabilities, observing advanced search features, user-friendliness in interface design, integration with other applications, etc. Finally, a number of characteristics have been selected that should be considered for future development, with the purpose of meeting different user's needs, namely portability, exportation and integration modes, additional structured content and documentation.

1 Italian Electronic dictionaries

An immense amount of research has been produced, dealing with electronic dictionaries' (ED) potentials, different ED typologies, lexical databases and computational lexicography. In this paper an attempt at evaluating state-of-the-art products will be conducted, aiming at highlighting strengths and weaknesses of Italian human readable electronic versions of printed dictionaries. Furthermore, we will try to illustrate some unexplored potentials emerging from user's specific needs in the exploitation of lexicographic data.

Six of the major Italian human readable electronic dictionaries (ED) have been observed, analyzed and evaluated: 1) *Garzanti* (2006) [GZ]; 2) De Mauro, *Gradit* (2003) [GRADIT]; 3) De Mauro, *Paravia* (2000) [DMP]; 4) Devoto-Oli (2005) [DO]; 5) Sabatini-Coletti (2006) [DISC]; 6) Zingarelli (2006) [ZNG]. Three points of interest have been compared, regarding retrieval power, performance evaluation and user-friendliness.

2 Retrieval system and query syntax

Information retrieval systems and query syntax are to be considered the main innovation in dictionary usage, introduced by EDs, compared to their printed versions. Retrieval in EDs implies a completely new way of searching linguistic material and a number of different new dictionary uses (Atkins 1998; Wiegand 1998; De Schryver, Joffe 2004). What will be debated here are not the actual capabilities in extracting significant linguistic information from the

lexicographic database, since it would imply a discussion on specific choices made by each lexicographer in marking up certain aspects, and applying certain rules for tag attribution. The object of this comparison will be retrieval power of the software associated with the specific information provided by the respective printed version. This means that retrieval power is an internal characteristic defined by the relationship between querying options and linguistic tags present in each printed dictionary. Strengths and weaknesses in look-up routes are thus to be evaluated within each single dictionary plan and design frame. Advanced search options let the user define multiple criteria in order to obtain not a single look-up entry, but a set of lexemes that satisfy those criteria. Table 1 shows different potentials of the six dictionaries at giving the user the possibility of elaborating formal patterns as a retrieval requisite.

	DISC	DMP	DO	GRADIT	GZ	ZNG
Headword	+	+	+	+	+	+
Inflected form	+	–	+	+	+	+
Complex words	+	+	–	+	+	–
Search in definitions	+	+	–	+	+	+
Logical operators	+	+	–	+	–	+
Wildcards	–	+	–	+	–	+
Beginning, containing, ending in	+	+	+	+	–	–

Table 1. Advanced Search Options

In this category fall the opportunity of starting a search from an inflected form (which are the headwords corresponding to the same textual token? E.g. this kind of search for *porta* leads to a couple of entries equivalent to *porta* as a singular noun, and *porta* as third person singular of the verb *portare*). The same system applies to complex words (idioms, collocations, etc.) and the option at exploring the entire dictionary as a linguistic corpus (including definitions). Advanced search options often include usage of logical operators (such as AND, OR, NOT, and in some cases NEAR, AT A DISTANCE OF, FOLLOWED BY, etc.) and, in a number of dictionaries, the ability to use wildcards (*.?) to specify formal patterns (e.g. all headwords and inflected forms starting with *psicologic-* and ending with only one additional character, thus excluding *psicologicamente*, etc.).

A relevant aspect connected with retrieval capabilities is linked to the possibility of further treatment of previously run searches, letting the user copy, print, save and export lists produced by advanced search (Table 2). From this point of view most products available are insufficient. Only GRADIT can save searches and create user defined dictionaries (made of a sub-set of the total entries selected by advanced search criteria). None of the dictionaries allows the user to copy, print or export list of words corresponding to a search, reducing drastically further treatment of lexicographic data (e.g. for learning tasks, working on lists for comparison and study, etc.).

	DISC	DMP	DO	GRADIT	GZ	ZNG
Copy single entries	+	+	-	+	+	+
Copy lists	-	-	-	-	-	-
Print single entries	+	+	+	+	+	+
Print lists	-	-	-	-	-	-
Export lists	-	-	-	-	-	-
Save search	-	-	-	+	-	-
User defined dictionaries	-	-	-	+	-	-

Table 2. Advanced search lists treatment potentials

Grammatical tags (N, V, Adj, etc.), inflectional paradigms, dependencies, frequency mark-up, usage tags (e.g. literary, rare, regional; technical, etc.), specific field domains (astronomy, medicine, computer science, etc.) can serve as criteria for advanced search in some of the observed EDs (Table 3).

	DISC	DMP	DO	GRADIT	GZ	ZNG
Grammar						
Grammatical category	+	+	+	+	-	+
Derivatives and compounds	-	-	-	+	-	-
Inflectional paradigms	-	-	-	+	-	-
Dependency and valence	+	-	+	-	-	+
Usage						
Frequent lexemes	-	+	+	+	+	-
Usage tags	+	+	+	+	-	-
Meaning usage tags	+	-	-	+	-	-
Register	+	-	+	-	-	-
Specific field domains	+	-	-	+	-	+

Table 3. Usage Mark-up

3 Performance evaluation

Among relevant aspects for ED usage in classroom and teaching activities is heaviness of software and hardware requirements (memory, space, OS, etc. see Table 4). The only Italian ED which does not require installation is Devoto-Oli (2005), although is nevertheless one of the slowest in charging the first search screen. Speed in performance generally decreases with more advanced search criteria selected and a wider lexicographic base (as in GRADIT, by far the larger and detailed). Highest retrieval speed is reached by DMP. Hard-disk space, which is a considerably significant parameter, varies from ca. 4 MB in Zingarelli to ca. 625 MB in GRADIT (due to the extensive lexicographic base) and Garzanti (due to audio files of pronunciation of all entries). In some cases there is no possibility of full installation, enabling the user to access ED content, without the cd-rom inserted in the drive: Devoto-Oli, GRADIT and Zingarelli do not offer this option. As to operating system requirements, only GRADIT and Devoto-Oli can be used on OS other than Microsoft Windows.

	DISC	DMP	DO	GRADIT	GZ	ZNG
Installation required	+	+	–	+	+	+
Hard-disk space (minimal)	204 MB	20 MB	–	15 MB	320 MB	4 MB
Hard-disk space (max)	204 MB	90 MB	–	624 MB	625 MB	4 MB
Full installation	+	+	–	–	+	–
RAM	128 MB	16 MB	non avail.	16 MB	128 MB	non avail.
OS	Windows	Windows	Windows Linux Macintosh	Windows Macintosh	Windows	Windows
Integration with Word Proc.	+	+	–	–	+	–

Table 4. System requirements

4 User-friendliness versus flexibility

Extremely relevant in ED evaluation is user-friendliness in the interface characteristics. User-friendliness is in itself a value, only when it does not affect overall performance and flexibility options. The more complex are, for example, advanced search options, more difficult is the creation of a friendly environment for the user, and the need of direct access to detailed documentation, at the price of a loss in immediacy of usage. User-friendliness is furthermore a relative concept, that can be applied to different user's needs and capabilities: learners', teachers', researchers' (Koren 1997; Iacobini 2003). Extremely hard is meeting all those need in one environment. Generally, Italian EDs correspond to the tendency in making prevail user-friendliness in those works which are poorer in search capabilities, like Garzanti and Devoto-Oli, and viceversa. GRADIT and Zingarelli, being the more flexible in advanced search options, are comparatively more complex in data and criteria selection methods. Besides the mere graphical interface (colours, screen layout and presentation, font choice and size), user friendliness can be mainly identified with the typology of selection of different criteria for advanced search. Items from a closed list (such as grammatical categories, usage marks, frequency data, chronology, specific field domain, etc.) can be selected and combined in various ways: choosing the item from a drop-down list, typing, clicking on a checkbox or calling a pop up list. Combination of different criteria, where possible, can be obtained by: logical operators (AND, OR, NOT selected from a list or typed in), +/– icons, checkbox symbols (), specific syntax (software specific wildcards, regular expressions). Specific syntax being the least usable and immediate combination method. Combination of criteria is sometimes only possible for different types of category (you can select all nouns which are rare and specific of astronomy, but you cannot select with one query all nouns *and* verbs which are rare *or* frequent and specific of any field *except* physics): as in the limited search capabilities of Devoto-Oli and Garzanti.

	DMP	DO	DISC	GZ	GRADIT	ZNG
Item selection	+	+	+	–	+	+
– drop-down list	+	+	+	–	+	–
– pop-up list	–	–	–	–	–	+
– typing	–	–	–	–	+	+
– checkbox	–	–	+	–	–	–

Combination of criteria	+	+	+	–	+	+
– drop-down list	+	+	+	–	+	–
– logical operators	+	–	–	–	+	+
– icons or symbols	+	–	–	–	–	–
– checkbox symbols	–	–	+	–	–	–
– specific syntax	–	–	–	–	+	+
Combination of the same type of criteria	+	–	+	–	+	+

Table 5. Selection and combination of search criteria

5 User's needs

The perfect ED being an idealized product for an idealized user, the integration of extreme flexibility demanded by the expert user (teacher, researcher and curious reader) and easy and friendliness of use asked by the student and young user is not always possible to accomplish. No publisher has developed the idea of producing both a student and a researcher version of the ED including the same content with different interface, search and processing options. Common to both types of user is need of portability (by full install), integration of ED resources with word processing and other textual environments (see Table 4). Common users also demand more standardized integration (as the inclusion of word and internet explorer toolbars), and additional contents like tables and summary points in hypertext form (as ordinary printed dictionaries traditionally present).

The advanced users' needs focus more on still lacking capabilities: mainly the ability to create user defined dictionaries (from queries and multiple queries), to export lists (corresponding to queries) in various forms (textual list of headwords) and export databases from queries (e.g. tables including criteria selected in the queries treatable by spreadsheets). The ability to use advanced search output as a first move towards teaching and developing learning aids, and researching vocabulary is the capital step in actual dictionary exploitation. EDs are not the same product as printed dictionaries, even if they might (and do) include the same content. Extraction and retrieval procedures enable new unpredictable and open-ended search tasks, that are wasted if not fully processable by other means (including word processing, and lexical bases for natural language processing).

Along with those critical demands, a further need for more explicitness in the lexicographer's choices specified in the documentation (that should be fully available in the cd-rom and in the installed directory) is clearly sensed. Only GRADIT offers both (De Mauro 1999a; 1999b; 2005). Explicit highlighting of mark-up application and theoretical and applied issues are extremely useful for the interpretation of lists retrieved and for the general understanding of the electronic resource power. What is formally the same query (e.g. selection for all nouns, in the most frequent words list, starting with letter <T>), does not produce the same results when administered to different EDs. To be able to appreciate and evaluate those differences a detailed account of the lexicographers' choices is indispensable (both in printed and electronic versions).

Publishers' preoccupation for software piracy should not stop suggested improvements (particularly full install features and export capabilities). All the six Italian EDs analyzed are

shared in peer-to-peer nets (like E-mule, Limewire, WinMX, etc.), most have cracked versions and serial key-generators widely available on the internet. All versions are fully working. For every system developed to prevent piracy there is a hacker that in a few minutes brakes the code and shares his work with others. For example, compelling the user to insert the cd-rom in the drive to use it is extremely disturbing, and does not prevent any hacker to provide a cd-rom image that can be run virtually (with DaemonTools or any other drive-emulation software), thus bypassing the problem. It does on the other hand prevent many buyers from not purchasing a product that is not fully accessible and portable.

EDs are exceptionally powerful tools. Those capabilities should be brought to their maximum of exploitation, by focusing on different users and their respective needs, and letting the ED as a linguistic resource be finally integrated with natural language processing, research demands, teaching and learning activities.

References

A. Dictionaries

A.a.V.v. (2006), Dizionario Garzanti di Italiano. Milano, Garzanti. (GZ)

De Mauro, T. (2003), Grande dizionario italiano dell'uso. Torino, UTET 1999-2000. (GRADIT)

De Mauro, T. (2000), Dizionario della lingua italiana. Torino, Paravia. (DMP)

Devoto, G. e Oli, C. (2005), Il Dizionario della lingua italiana. Firenze, Le Monnier. (DO)

Sabatini, F. e Coletti, V. (2006), DISC – Dizionario della Lingua Italiana. Milano, Rizzoli Larousse. (DISC)

Zingarelli, N.(2005), Lo Zingarelli 2006: Vocabolario della lingua italiana. Milano, Zanichelli. (ZNG)

B. Other Literature

De Schryver, G-M. (2003), 'Lexicographers' Dreams in the Electronic-Dictionary Age', *International Journal of Lexicography* 16/2, pp. 143-199.

De Schryver, G-M. and Joffe, D. (2004), 'On how electronic dictionaries are really used', *Proceedings of the 11th EURALEX international congress* held in Lorient, France, July 2004. Université de Bretagne Sud.

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

Wiegand, H. E. (1998), *Wörterbuchforschung: Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie.* 1. Teilband. Berlin, Walter de Gruyter.

Koren, S. (1997), 'Quality versus Convenience: Comparison of Modern Dictionaries from the Researcher's, Teacher's and Learner's Points of View', *TESL-EJ*, 2 (3), A-2.

De Mauro, T. (1999a), 'Introduzione' in De Mauro, T. (1999-2000) *Grande dizionario italiano dell'uso.* Torino, UTET, vol. I, VII-XLII.

De Mauro, T. (1999b), 'Postfazione' in De Mauro, T. (1999-2000) *Grande dizionario italiano dell'uso.* Torino, UTET, vol. VI, pp. 1163-1183.

De Mauro, T. (2005), *La fabbrica delle parole.* Torino, UTET.

Iacobini C., (2003), 'I dizionari della lingua italiana su cd-rom come strumento per la ricerca linguistica', in Rainer, F., Stein, A.(a cura di), *I nuovi media come strumento per la ricerca linguistica*, Frankfurt, Peter Lang, pp. 61-73.