To appear in in *Literary and Linguistic Computing*, 16(1): 123–139, 2001.

Kirrkirr: Software for browsing and visual exploration of a structured Warlpiri dictionary

Christopher D. Manning,[1] Kevin Jansz,[2] and Nitin Indurkhya[3]
[1]Stanford University, USA
[2]University of Sydney, Australia
[3]Nanyang Technological University, Singapore

Correspondence:
Christopher Manning
Department of Computer Science, Gates 4A
353 Serra Mall
Stanford CA 94305-9040
USA
manning@cs.stanford.edu

Abstract

This paper presents an overview of the goals, architecture, and usability of Kirrkirr, a Java-based visualization tool for XML dictionaries, currently being used with a dictionary for Warlpiri, an Australian Aboriginal language. It discusses the underlying lexicon structure, and shows how a computer interface can effectively select from and display that content in various ways. The views of the dictionary include a graph view, which shows a network of semantically related words, and a formatted text view that can be customized via XSLT stylesheets. The paper argues that indigenous language dictionaries have normally been written for linguists, while the educational needs of other users have not been adequately met. It discusses the strengths of a computer dictionary interface in providing more help to native speaker users than a conventional dictionary, but argues that the possibilities for the visualization of dictionary information on computers have so far been insufficiently exploited. The paper concludes by briefly discussing observational and task-based testing of the dictionary with native speakers and learners.

This paper discusses the goals, architecture, and usability of Kirrkirr, a Java-based visualization, search, and browsing tool for XML dictionaries.[1] In particular, we discuss its use with a dictionary for Warlpiri, an Australian Aboriginal language, focusing on its potential for providing practical, educationally useful dictionary access for endangered languages, at a reasonable labor cost.

[1] We thank many people for their help. The work presented here would have been completely impossible without the assemblage of richly structured Warlpiri lexical materials over many years by Ken Hale, Mary Laughren, Robert Hoogenraad, and many other *kardiya* and *yapa.* We hope that our work is of some assistance to them in return. We thank Mary Laughren for access to the Warlpiri Dictionary; Jane Simpson for her central role in the project of which this work is a part; Miriam Corris, Jane Simpson, David Nash, and Ben Hutchinson for work in testing Kirrkirr; Robert Hoogenraad, Jenny Green, Carmel O'Shannessy, and Margaret Carew for arranging and facilitating much of the work of the testers; and audiences at the Workshop on Linguistic Exploration accompanying the 2000 LSA Annual Meeting, ALLC/ACH 2000, and in a course taught by Manning and Simpson at the 2000 Australian Linguistic Institute in Melbourne. This work was supported in part by Australian Research Council Small and Large Grants to Simpson and Manning in 1998, 1999, and 2000–02.

Introduction

Dictionaries on Computers

While dictionaries on computers (CD-ROMs or on the web) are now common, there has been surprisingly little work on innovative ways of utilizing the capabilities of computers for visualization, customizable hypertext, and multimedia in order to provide a richer experience of dictionary content. The displayed dictionary entries normally attempt to mimic the layout of conventional paper dictionaries, except for limited attempts at providing hypertext linking, but look worse because of the much lower display resolution, and often imperfect reproduction of fonts, etc. Moreover, most electronic dictionaries present the search-dominated interface of classic information retrieval (IR) systems: a box in which to enter the search word. This is only effective when the user has a clearly specified information need and a good understanding of the content being searched. The ability to browse often makes paper dictionaries easier and more pleasant to use than such electronic dictionaries. Search interfaces are ineffective for information needs such as exploring a concept. Some work in IR has emphasized the need for new methods of information access and visualization for browsing document collections (e.g. Pirolli et al. 1996), and we wish to extend such ideas into the domain of dictionaries.

Beneath the surface as well, the internal structure of most current Machine Readable Dictionaries (MRDs) merely mimics the structure of the printed form from which they are derived (Boguraev 1990). Although there has been some work, notably WordNet (Miller et al. 1993), which has involved a fundamental rethinking of dictionary content and organization, it has not had much impact on dictionary users (though see Schechter 1997).

While reference books differ also in the quantity of information they contain about topics (a book on birds has more information on birds than a standard dictionary), many of the well-established forms of reference books differ primarily in the means that they provide for *indexing* their content. A dictionary indexes its content by alphabetical order. A thesaurus indexes its content by concepts, so words of similar meaning are easily found. Some pictorial dictionaries index material into terminology sets, such as words that are used for referring to equipment used to play cricket, and perhaps verbs that are involved in the play of the game. A field guide to birds typically indexes the information primarily by attributes of the color, shape, and perhaps biological family of birds. In a computer implementation, one could hope to have a rich warehouse of lexical information, where all these means of indexing and linkage are available to the user, but this hope has yet to be adequately realized.

A particular interest of ours is dictionaries for minority languages. Here economic, motivational, and support reasons all point to an important role for computers. We discuss the economic issues here, and the other issues in the next subsection. For minority languages, there will usually not be the monetary resources, and commonly also not the human resources, for producing many different dictionaries (corresponding to the learner's, foreign learner's, concise, comprehensive, etc. dictionaries available for major languages). There is thus much greater need for clever use of computers in automatically selecting and formatting lexicographic content to meet user needs without further editorial intervention. For indigenous languages, we find that dictionary structure and usability have usually been dictated by professional linguists, who see their primary task as documenting the language for other language professionals, while the needs of others (speakers, semi-speakers, young users, second language learners) are not met (e.g., see Goddard and Thieberger (1997) for general background on dictionaries for Australian Aboriginal languages). There is a clear parallel with Weiner's (1994) remarks on the Oxford English dictionary that the initial

purpose was “to create a record of vocabulary so that English literature could be understood by all. But English scholarship grew up and lexicography grew with it … inevitably parting company with the man in the street”. Our goal is to avoid this by exploring fun computer dictionary tools that are effective for language learning, browsing, and research by various communities of users.

Electronic Dictionaries for Language Learners

So far there has been little use of MRDs in education. For instance, Kegl (1995) writes: “Originally, this paper was intended as a survey of educational applications using MRDs. As far as I have been able to determine, no such applications currently exist.” Further, she goes on to discuss how standard dictionaries are reference works, ill-suited for use as learning tools, and discusses how studies of American “dictionary skills training” course modules show that many tasks achieve little educational benefit (though they do presumably teach word lookup). However, if we look more broadly, lexical information clearly *is* educationally useful. Standard high school language texts are full of processed lexical information, such as vocabulary lists or terminology sets accompanying a topical chapter like “At the seaside”, pictures with parts or entities named, and notes on the usage of words. What we hope to show is that by selectively taking and linking material from a rich lexical database, we can provide a fun, and educationally useful tool. A lexicon is not just a list of words, but a vast network of associations between words and across the concepts represented by words. Traditional paper dictionaries offer very limited ways for making such networks visible. A central aim of the project is to give people a better understanding of this conceptual map, and computer tools seemed a useful way of achieving this.

There has been relatively little work on the visualization of dictionary information, especially for end users (though see Munzner et al. 1999). For example, thinking of the lexicon as a semantic network with various kinds of links was a leading idea of the WordNet project (Miller et al. 1993), but while this organization fundamentally departed from the paper dictionary tradition, the database came with a simple, search-oriented textual interface that failed to do justice to the richness of the underlying data. Some projects have then taken the WordNet data and represented it in visual displays, for example plumbdesign’s Visual Thesaurus (plumbdesign 1998). However, most of the work has provided either too sophisticated or quite busy and potentially confusing interfaces. A particular goal of our work was to provide a simple and clean, yet appealing and interactive, graphical layout of words that was appropriate to our potential users, including in particular children.

Parallel work in the same project as the work reported here has studied the use and usability of paper dictionaries for endangered languages, in part hoping to provide insights that are relevant to the design of computer-based dictionary versions. Corris et al. (2000a, 2000b, forthcoming) report that limited proficiency in literacy and in other dictionary skills (alphabetical order, conventions, and abbreviations) greatly limited the utility of paper dictionaries for Australian Aboriginal users. Users lost their place, became confused by the overcrowding of information, misunderstood or failed to understand conventions and notations like subentry structure, parts of speech and abbreviations labeling cross-references. People regularly took several minutes to complete fairly simple word lookup tasks (e.g., looking up a definition or a synonym for insertion into a crossword). Our hope is that many of these problems can be mitigated in a computer interface, for reasons ranging from the less pressing space restrictions to the ability to provide more in the way of learner supports.

A particular goal has been to design an interface usable by, and interesting to, young users and others acquiring literacy or the Warlpiri language. From this viewpoint, the low level of literacy in the region, and the inherently captivating nature of computers (Brown

1985) suggest that an e-dictionary is potentially more useful than a paper edition. Among other benefits, we can provide an interface less dependent on good knowledge of spelling and alphabetical order. Wallace et al. (1998) note that "information seeking is a complex process which is often not attended to in K-12 [i.e., primary and secondary] education". A computer interface can be adaptable in catering to different needs by providing multiple entry points, and various kinds of learner supports. Additionally, it can support active reading via note-taking, and other forms of interaction.

Architecture of Kirrkirr

Our goal has been to provide a fun dictionary tool that is effective for browsing and incidental language learning, as well as focused information finding, by users of different ages and abilities. Kirrkirr is our prototype design that attempts to achieve some of these goals (Jansz et al. 1999). The interface attempts to more fully utilize graphical interfaces, hypertext, multimedia, and different ways of indexing and accessing information. In particular we attempt to address Sharpe's (1995) "distinction between information gained and knowledge sought". The speed of information retrieval that e-dictionaries deliver, and the focused decontextualized search results they provide, can frequently lead to loss of the memory retention benefits and chances for random learning that manually searching through paper dictionaries provides.

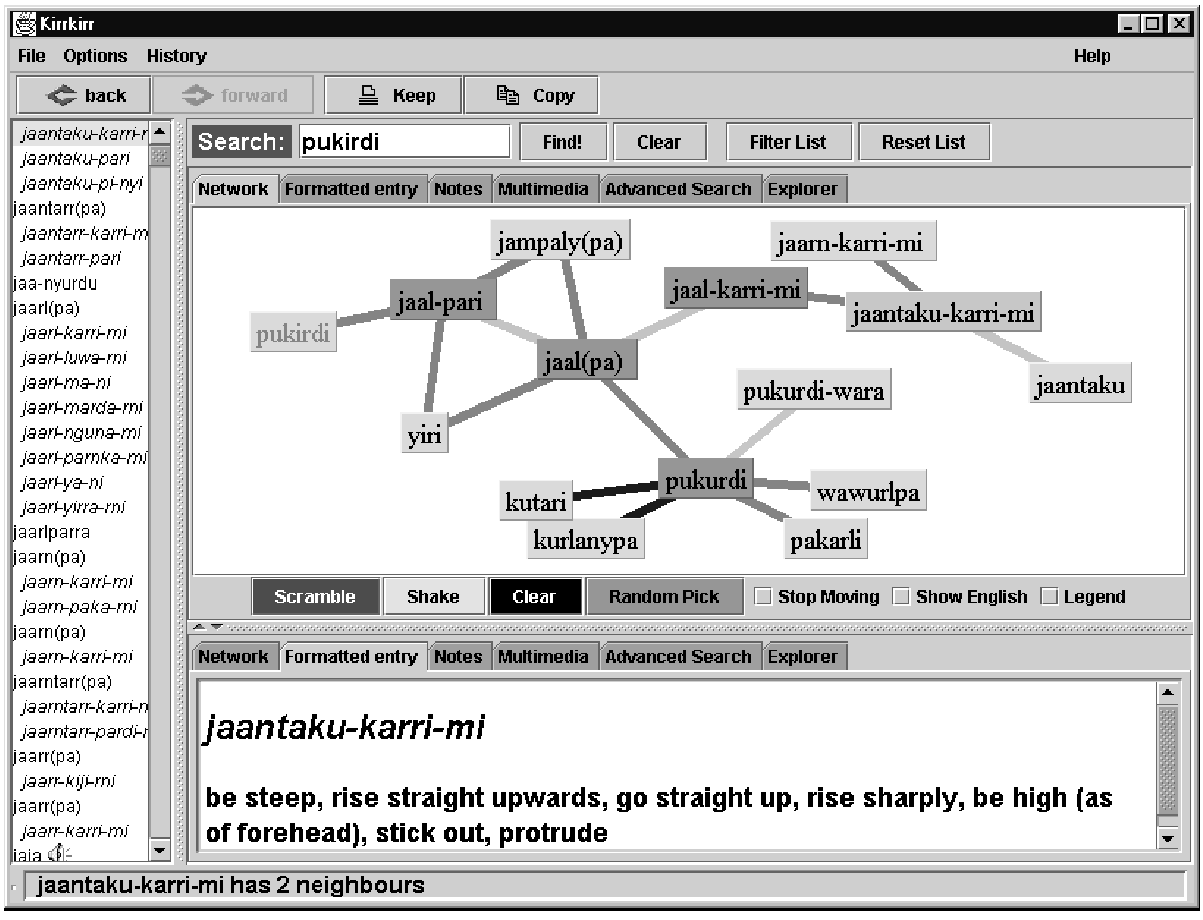

Fig. 1 A view of the Kirrkirr interface

One screen shot of Kirrkirr is shown in Fig. 1. The system is written in Java and runs on all major platforms (including Windows, Mac, Unix). Down the left-hand side, the program shows a listing of headwords. While accessing something the size of a dictionary via a scroll-list is completely impractical, this interface element is important in giving concreteness to the dictionary. Compared with physical books, computer interfaces often lack tangibility: the user has no idea what lies among the electrons beyond the screen. The scroll-list immediately makes it clear that we have an organized list of words. A first-time user can start by clicking on one of these words, rather than having to be able to spell a word correctly. This fits with one of our major design goals: the interface should at all times *show* people words, but knowing words, their spelling, and concepts like alphabetical order should not be required in order to start using the interface. As words are searched for or selected by other means, the word list scrolls to the appropriate position, so that at all times the user can see surrounding words, just as in a paper dictionary. Moreover, this commonly means that a user need only type a few letters and can then look at the list – modeling a traditional benefit of paper dictionary usage. The top of the right-hand side of the interface provides a conventional search box. The rest of the screen can display one or two panes that give other views of the dictionary, or of an individual word, or an advanced search interface. These are selected via tabbed panes.

Lexical structure and XML

Although the design of our dictionary interface is general, we initially targeted Warlpiri, a language of Central Australia, for which there has been an extensive on-going project for the compilation of semantically rich lexical materials (Laughren and Nash 1983, Laughren et al. forthcoming).[2] Linguistic data collection began in the late 1950s, and computer-based dictionary compilation began in the early 1980s (some of the early history appears in Laughren and Nash 1983). Due to Ken Hale's presence at MIT, the project witnessed many of the trends and epochs in computing (early timesharing systems and laser printers, a parser written in Lisp, etc.). The result is the richest compilation of lexical material for an Australian language (there are about 10,000 headwords, with English and often vernacular definitions, and extensive exemplification, cross-referencing, and dialect information). However, until now this information has not been made available to the community other than as a fairly raw printout of marked up text.

The editors of the Warlpiri dictionary (Laughren et al. forthcoming) maintain their data in plain text files in a non-standard marked-up format, in which elements are coded with backslash-initial tags. The use of backslashes makes the files look superficially similar to SIL Standard Format (Coward and Grimes 1995) but the tags are ultimately derived from Runoff (an early typesetting system) commands, and the provision of end tags and element nesting makes the representational power similar to SGML. We converted this data into a richly structured XML version (XML 1998), using a stack-based (Document Object Model tree walking) parser written in Perl. The parser attempted to correct – the quite numerous – structural errors in the input dictionary (normally by adding or deleting tags as needed). Regardless of the editorial correctness of the changes it made, the parser could guarantee that the resulting XML was not only well-formed but valid according to the Warlpiri dictionary

[2] Warlpiri is a morphologically rich, morphologically ergative language, famous in the theoretical linguistic community for its non-configurational structure (Hale 1983, Simpson 1991). It is spoken by about 3000 people. It is the first language of all three Warlpiri communities we discuss, and there are active bilingual programs (although they are under threat from politicians). Many people can write Warlpiri as well as English, though most old people are illiterate. See Corris et al. (forthcoming) for further information.

DTD. We show elements of this DTD in Fig. 2. The DTD is fairly loose, and except for a few encoding decisions and augmentations for additional information types, it represents a fairly straightforward translation of the structure of the original source files. The dictionary is a list of entries. Entries contain many types of information after the headword, including separate gloss (GL) and definition (DEF) fields, examples, dialect and register information, and many types of links including synonyms, antonyms, alternative forms, and lists of preverbs that combine with a verb (PVL). Many entries are organized into senses (SENSE), and/or into paradigm examples (PDX). This is a distinctive feature of the Warlpiri dictionary whereby senses are used when the lexicographers think that the word has multiple meanings in Warlpiri, but paradigm examples are used when they think that there is only one underlying Warlpiri meaning, but that it occurs in different contexts, where it would be glossed differently by English speakers (for instance, a word defined as "having the quality of not causing unpleasantness" has paradigm examples where it is glossed as "well-behaved" when applying to humans, and "mild" when applying to food or drink). See Laughren and Nash (1983) for further discussion of lexicographic decisions underlying the Warlpiri dictionary. The dictionary recognizes headwords that are homophones, indicated by a number attribute of the headword (HW) element, and makes use of subentries for derived forms. In our XML translation, subentries are promoted into full entries, but the subentry-main entry linkage is shown via two further cross-reference types (SE and CME). This reflects Corris et al.'s (forthcoming) dictionary usability results where subentries generally caused confusion. Testing of Kirrkirr suggests that turning this relationship into another form of crossreferencing makes things easier (at least given the visualization methods discussed below). Except for a DICTIONARY being a list of ENTRY elements, we make no particular effort to use the dictionary DTD of the text encoding initiative (Sperberg-McQueen and Burnard 1994). While we saw great value in the use of XML, we saw no particular value in trying to map the Warlpiri dictionary to an existing DTD, particularly as the Warlpiri dictionary editors continue to edit the dictionary in the original format. It is thus simpler to use a loose DTD that fairly straightforwardly encodes the existing structure of the Warlpiri dictionary.

```
<!ELEMENT DICTIONARY (ENTRY+)>
<!ELEMENT ENTRY (HW, FREQ, IMAGE?, SOUND?, (POS | DOMAIN | GL | DEF |
DIALECTS | REGISTERS | EXAMPLES | CF | SYN | XME | ANT | ALT | PVL | CME | SE |
CSL | CMP | REF | CM | DERIV | CRITERION | SRC | REFA | LAT | REM | PDX |
SENSE)* )>
<!ELEMENT SENSE ( (DOMAIN | GL | DEF | DIALECTS | REGISTERS | EXAMPLES |
CF | SYN | XME | ANT | ALT | CM | CRITERION | CSL | PDX)* )>
<!ELEMENT PDX (CRITERION, (GL | DEF | EXAMPLES | CF)* )>
<!ELEMENT EXAMPLES (EXAMPLE+)>
<!ELEMENT EXAMPLE (WE, ET)>
```

Fig. 2 Elements of the DTD for the Warlpiri dictionary

The entire dictionary is stored as one large (10Mb) XML file. All standard XML parsers of which we are aware attempt to parse an entire XML file into memory, but for a file of this size, the space and time requirements make this highly unwelcome, so we have modified XML parsers to only parse a single entry as needed. We at present use custom (ad hoc) indexing of the XML file to provide quick access to appropriate entries on lookup. An experimental version of Kirrkirr that makes use of XQL (XQL 1999) via use of the GMD-

IPSI XQL engine (GMD-IPSI 1999) is described in Jansz et al. (2000). However, at present the ad hoc indexing is faster (if ultimately less flexible), and the absence of an established standard query language for XML limits the value of investing effort in this alternative approach. Nevertheless, we hope to move to access via a standard query language, when such facilities exist and are efficient.

For dictionaries with plain textual content behind them, there is little that they can provide in the way of output but an on-line reflection of a printed page. In contrast, XML allows definition of the precise semantics of the dictionary content (which includes a large amount of defining and encyclopedic information, and exemplification and usage of a word and its associations), while leaving unspecified its form of presentation to the user. In the following sections we show how we exploit this flexibility, by having the program mediate between the lexical data and the user. The interface can select from and choose how to present information, in ways customized to a user's preferences and abilities.

Visualization

As we noted at the beginning, a lexicon contains many networks of associations between words, and users are often interested in exploring these connections, yet traditional dictionaries have quite limited means for making these comprehensible to users. The goal of the Network pane of Kirrkirr is to show these connections in a brightly colored (and animated) display. As shown in Fig. 1, Kirrkirr shows a color-coded network of semantic links between words, which can be explored, manipulated and customized interactively by the user (Jansz et al. 1999). What is needed for this is essentially a good solution to the classic graph-layout problem. At present we are using the animated graph-drawing techniques of (Eades et al. 1998, Huang et al. 1998). In their spring algorithm, the words in a semantic network become nodes which are held apart by gravitational repulsion, but kept from becoming too far apart by springs which have a natural length. This graph algorithm differs from most others by providing iterative updating of the graph layout, which means that users can drag nodes across the screen, and the algorithm will cause other nodes to flee out of the way, while words related to another word are dragged along. The software maintains a set of focus nodes to prevent overcrowding: as the display fills up, less recently used nodes are automatically deleted.

The detailed semantic markup of the dictionary, with many kinds of semantic links (such as synonyms, antonyms, hyponyms, and other forms of relationships) allows us to provide a rich browsing experience. For example, the ability to display different link types graphically as different colors solves one of the recurring problems of the present world wide web, with its one type of link: users have some idea of what type of relationship there is to another word before clicking. Consistent color-coding of text and edges between the graphical and formatted text displays is used to reinforce the meaning of the link types, and a legend can also be displayed.

To augment the traditional semantic relations provided in the dictionary, we have also experimented with additionally providing linkages derived automatically from collocational analysis of Warlpiri text. This works with some success but is also an area that requires further thought and research. On the one hand, while the amount of online Warlpiri text is quite large for an indigenous language (around two hundred thousand words), this is tiny by the standards of corpus linguistics, and this makes it difficult to impossible to reliably find collocations for many words. There are also editorial decisions about the validity of the derived links. While the core dictionary is a carefully human-edited set of definitions and cross-referencing, it is now being augmented with automatically generated, unchecked

material. Since people normally expect everything in dictionaries to be correct and checked, we decided to have collocational links turned off by default in the Kirrkirr interface.

Formatted Entries: Dictionary customization through XSL stylesheets

Complementing the use of tree and network visualization and other media, users of a dictionary still necessarily require a conventional textual representation of dictionary entries. However, quite different textual displays are appropriate for different user groups. For instance, items such as abbreviations for parts of speech, and other grammatical notes, detailed decompositional definitions, or simply long dense entries can be confusing for most Aboriginal users (Corris et al. 2000b, forthcoming). In Kirrkirr, formatted dictionary entries, displayed using HTML, are produced at runtime from the underlying XML dictionary as needed by the use of XSLT stylesheets. XSLT is a declarative language for transforming a structured XML document by applying rules to different element types (XSLT 1999). This allows us to provide different subsets of the information in the dictionary database, differently formatted, as is appropriate for different users and user needs. Three different renderings of the same dictionary entry are shown in Fig. 3. The default rendering gives little more than a word and some glosses in large, easy-to-read type. This is suitable for beginning users with simple needs about what words mean, and resembles a word list, or beginner's dictionary. The full entry format shows much of the information in the dictionary database (while still omitting information about sources, frequencies, comparative information, and certain grammatical notes). As indicated by the scrollbar, only part of the entry is shown in the figure. We note also that many entries are much more complex than the one shown in the figure, with organization into various senses, or via usage into paradigm examples. These entries provide extensive exemplification of word usage, and conventional hypertext for navigating between entries, via related words. We provide a color-coding of different kinds of semantic relationships between words, which is consistent with that in the network display, so as to maximize learning and transfer.

In testing an earlier version of the dictionary in which something similar to the current full entry format was the default textual display, we noticed that some users with good Warlpiri literacy would consistently read Warlpiri definitions and examples in preference to looking at the English (Corris et al. forthcoming). In response, we now also provide a rendering in which Warlpiri is used in preference to English, as is also shown in Fig. 3. Many entries in the Warlpiri dictionary have definitions in Warlpiri, and thus we can provide a monolingual dictionary entry. Others do not, and then the system falls back to using English glossing supplemented by Warlpiri examples. For non-Warlpiri speaking readers of this article, note that the Warlpiri definition shown in the figure also appears as an example, with a translation, in the full dictionary entry.[3]

In total there are at present five such predefined renderings (the other two are one that includes the glosses and the cross-references, but not the examples, and one that shows everything in the dictionary entry). Further work remains to be done to determine if our current renderings should be augmented or changed to better suit user abilities and needs.

[3] Most of the Warlpiri definitions in the dictionary, as supplied by the first generation of Warlpiri lexicographers, have a character similar to this one, making extensive use of ostension, exemplification, and word relationships. It remains to be seen whether future generations will move to something closer to the Aristotelian notions of necessary and sufficient conditions more common in European dictionary definitions.

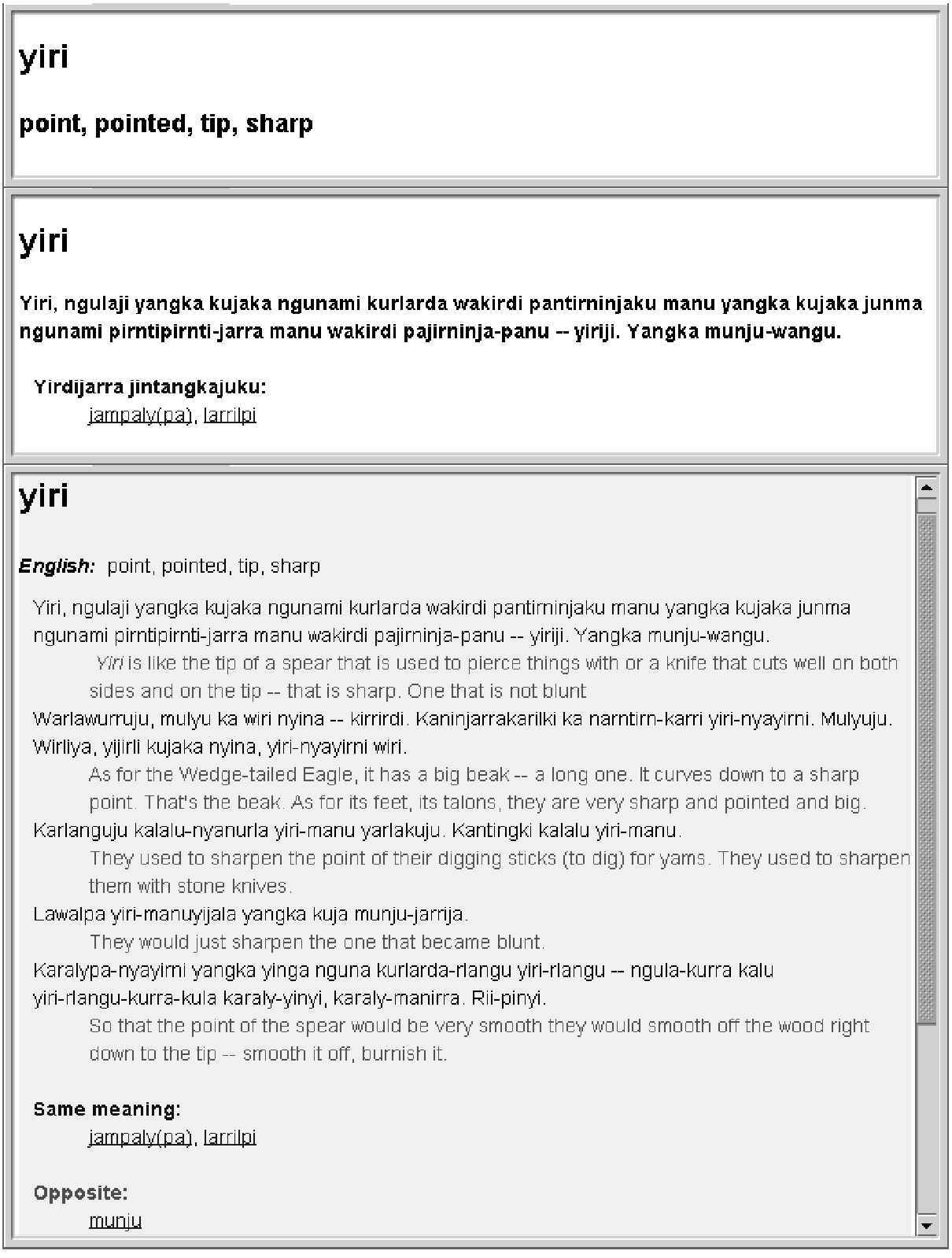

Fig. 3 Three different renditions of a dictionary entry achieved via XSLT

Other components

In addition to the above, the dictionary incorporates various other components – and some further components are planned. Alphabetical order is just one indexing strategy; users also frequently want to find words via semantic domains, such as words for trees, or games. The Warlpiri dictionary incorporates a rich classification into semantic domains, which at lower levels is organized via Warlpiri ethnographic classifications, and Kirrkirr makes these available via a standard tree-structured view. This provides a hierarchical view of the words via semantic domains, and is potentially particularly useful in various educational activities – teachers frequently want lists of reptiles, or trees, etc.

The program provides standard multimedia enhancements – the user can hear words and see appropriate pictures. These simple features are extremely important. In Corris et al.'s (2000b, forthcoming) studies users with low levels of literacy are frequently frustrated by paper dictionaries when, on having finally successfully looked up a word that they don't know in the language, they at the end still don't know how to pronounce it because they are unaware of such things as conventions for marking stress (if indeed such things are marked in the dictionary). The multimedia and semantic domain views of the dictionary are shown in Fig. 4.

The dictionary also provides a more sophisticated search interface, with a user-friendly console where search results can be sorted and manipulated. As well as standard keyword search, which can optionally be restricted to appearance within a specified XML element (such as Warlpiri headwords, or English glosses), the system provides two features targeted towards two principal groups of users. Linguists often want to search for particular sound patterns (such as certain types of consonant clusters), and so the system allows regular expression matching for such expert users.

On the other hand, the limited literacy level of many potential users means that they will have problems looking up words. In part this is due to particular problems whereby the phonetic orthography of Warlpiri does not match very closely to the (rather arcane) spelling rules of English in which their literacy skills are commonly based. To alleviate this problem, we have implemented a "fuzzy spelling" algorithm which attempts to find the intended word by using rules which capture common mistakes, sound confusions and alternative spellings. Fuzzy spelling is implemented on the fly by using regular expressions in searching (rather than a SOUNDEX strategy where sounds or letters are mapped onto equivalence classes). This nevertheless achieves satisfactory performance with modern hardware. The fuzzy spelling algorithm is also used behind the scenes with the simple search box, if exact lookup fails.

Finally, there is a Notes pane. While the dictionary is primarily a read-only browsing and search interface (and not a dictionary editing tool), this allows users to annotate, correct or personalize dictionary entries, and encourages active learning. In the future we hope to do more to encourage activity, such as by incorporating exercises and tasks into the dictionary – several teachers have suggested to us that this would be valuable.

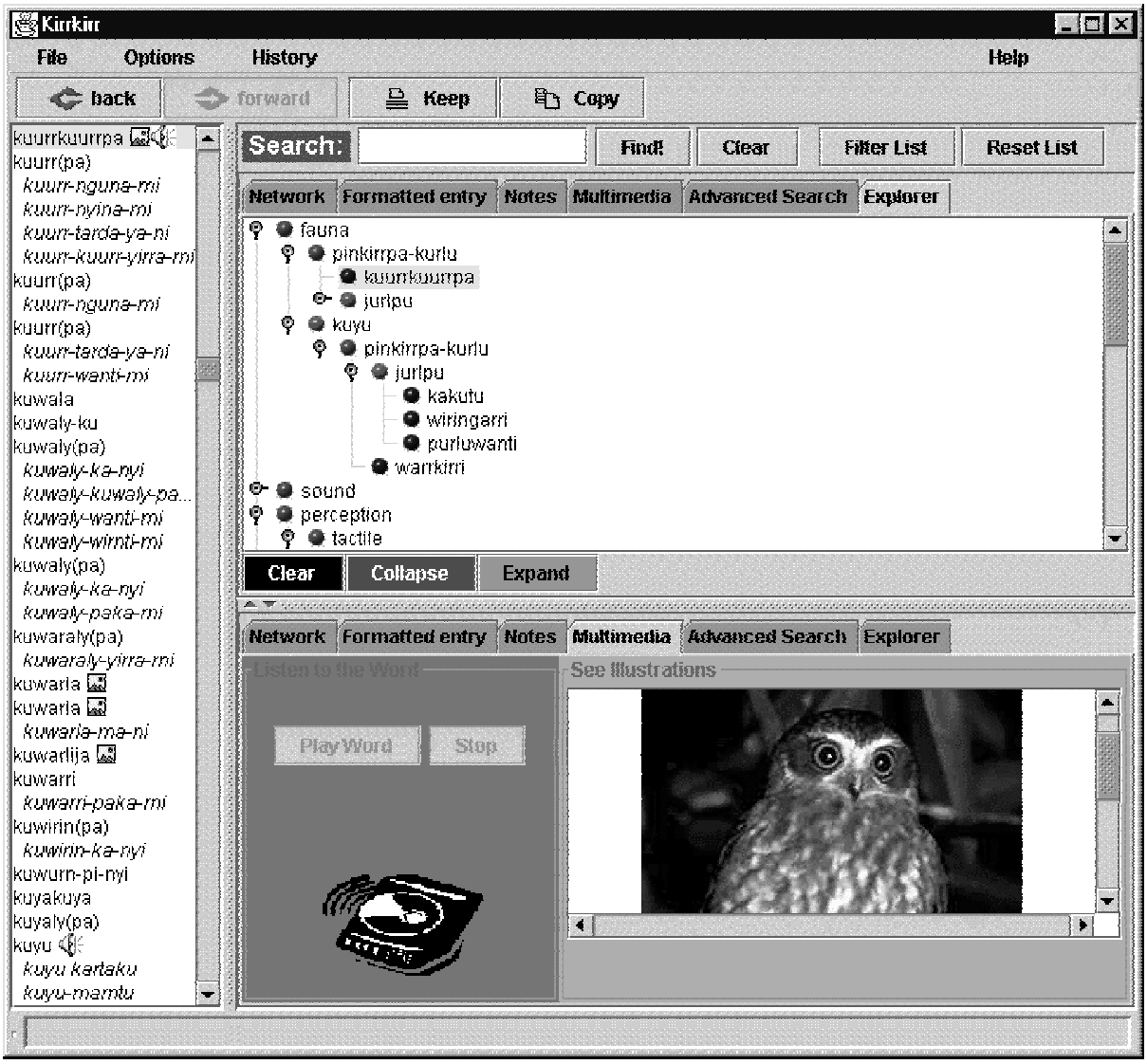

Fig. 4 The semantic hierarchy and multimedia panels of Kirrkirr

User studies and evaluation

Project members have done demonstrations and studies of the usefulness and usability of several versions of the software. An initial version of Kirrkirr was trialed by Mim Corris through visits to Yuendumu and Willowra in February 1999, and by Jane Simpson in visits to Lajamanu and Yuendumu in May 1999. This motivated some of the further development of the software, and a somewhat improved version was demonstrated in February 2000 by Ben Hutchinson in Alice Springs and at Batchelor College, and extensively tested by Jane Simpson in teaching literacy courses and other activities at Batchelor College and at Lajamanu in June 2000. Lajamanu, Yuendumu, and Willowra are three of the main Warlpiri centers, where bilingual education has been available, and Batchelor College provides training in indigenous languages and literacy. These trials have involved completing dictionary tasks, and observational use with primary and lower secondary students and

trainee Warlpiri literacy workers, and comments from teachers and other adults. The results and details on the methods and materials appear in full in Corris et al. (2000b, forthcoming), but we will briefly make a few observations and mention a few anecdotes here.

Among the difficulties with the study is that potential users often have little or no prior experience with dictionaries, particularly not with Warlpiri dictionaries, and thus have no standard for comparison. Moreover, the general circumstances make it difficult to imagine formal tests of effectiveness, and most of our results come from qualitative data of students working through dictionary task sheets together with a researcher.

Some teachers have initially expressed the view that the interface is too advanced for children, and might be more useful as a resource for teachers. However, in our experience, teachers underestimate the facility of today's young (indigenous) children with computers. For instance, one 10 year-old child (admittedly one of the brightest kids) returned after a demonstration of Kirrkirr and played with it for 2 1/2 hours. In such cases, opportunistic exploration of networks of related words seems to be able to capture and maintain attention. As another simple indication of computer knowledge, the first version we trialed lacked a Back button (akin to the one on a web browser), and primary school students were quick to point out this deficiency.

For younger children, clicking on things and making them move sometimes clearly had more interest than details of meaning, but older primary and post-primary children were quite thoughtful in their dictionary observations, and able to use Kirrkirr to complete dictionary tasks. Project members have tried to compare people using Kirrkirr versus paper dictionaries to complete such tasks, but the many confounding factors (prior knowledge of computers, or in particular of Kirrkirr, how well laid out the paper dictionary is, user's typing ability, and eyesight, or just the presence of random interruptions) make it difficult to give meaningful quantitative results.

People liked and enjoyed following and explaining synonym links, alternate and dialectal forms, and links to derived words in subentries. On the other hand, the semantically imprecise general "cf" links were confusing, as were (somewhat more surprisingly) antonym links. The concept of an antonym did not appear to be immediately apparent to users.

In general reactions have been quite enthusiastic. The dictionary does appear to succeed in creating and maintaining interest, and it meets with some success in catering to different user's needs. Young children can click around as in a game, and often go in search of pictures, while older users are commonly much more interested in the extensive exemplification of Warlpiri usage. As noted above, some users primarily use Warlpiri definitions and examples for information, whereas others, with stronger literacy in English, primarily turn to English. One way in which the latter group's needs are currently not met is that, even though one can search via English glosses, etc., the dictionary and the interface is primarily organized around Warlpiri lookup. An area of future work is to try to give greater equality to access from English, despite the fact that the underlying database is fundamentally a Warlpiri-English dictionary. We have received suggestions on how to make it a better basis for classroom activities, which we hope to incorporate in future versions. In the coming year (2001), we will have the dictionary installed permanently in schools (this has required some hardware investments – the dictionary requires a computer with 64 Mb of RAM and a minimum processing power equivalent to about a 200 MHz Pentium, and while this is basically the minimal configuration for computers that can be purchased in 2000, existing school computers are not always sufficient). This should give much better opportunities to judge the utility of Kirrkirr in natural use by experienced users. We conclude this section with one positive anecdotal report (and note that we've since fixed the formatted entry display bug!):

> One of the introductory Warlpiri literacy students, who had not been very interested in the literacy class, spent nearly 3/4 hour looking at Kirrkirr apparently in absorbed concentration. She wasn't especially interested in the sound and picture possibilities. She moved between words, scrolling along the list, typing in the search, clicking on the words in the network pane. She wasn't even put off when the dictionary definitions stopped appearing – looking at the networks of words instead. This is quite unlike her attitude to the backslash-coded electronic dictionary (where she lost interest quickly because of the difficulty for her of narrowing down searches). After the Kirrkirr demo she asked if she could have a printed dictionary to take away with her to use in camp to learn the words. I interpret this as a desire to learn words in her own time and place. (Jane Simpson, report May 1999)

Conclusions

Kirrkirr is a prototype of what one can do to develop new ways to organize and visualize lexicons. We have extended the state-of-the-art for electronic dictionaries in several ways. Our implementation develops a unified solution to the problems of storage, processing, visualization and learning. The construction of lexical databases and dictionaries has been taken a step further to benefit people such as language learners, who could truly gain from a better interface to dictionary information. Furthermore, we have tackled the problem of catering to a much more diverse range of potential users than encountered in typical studies of dictionary usability (e.g., Atkins and Varantola 1997), especially ones with low levels of literacy. We have addressed the challenge of making dictionary information accessible and usable by various levels of users through the creation of an application that mediates between well structured data and users' needs and insights in searching/browsing and presentation. Our system has attempted to reduce the importance of knowing the written form of the word before the application can be used, while having ample opportunities to learn written forms. Features such as an animated, clearly laid out network of words and their relationships, multimedia, and hypertext aim at making the system interesting and enjoyable to use. At the same time, features such as advanced search capabilities and note-taking make the system practical as a reference tool. Having designed the system to be highly customizable by the user, it is also highly extensible, allowing new modules to be incorporated with relative ease. We thus think that it is a good foundation for an electronic dictionary, and while the focus of this research has been on Warlpiri, this research (and the software constructed) can be easily applied to other languages.

References

Atkins, B.T.S. and Varantola, K. (1997). Monitoring dictionary use. *International Journal of Lexicography* 10: 1–45.

Boguraev, B. (1990). Data Models for Lexicon Acquisition. *International Workshop on Electronic Dictionaries,* pp. 70–86. Oiso, Kanagawa, Japan. EDR technical report TR-031.

Brown, J. S. (1985). Idea Amplifiers. *Educational Horizons*, 63: 108–112.

Corris, M., Manning, C., Poetsch, S., and Simpson, J. (2000a). Dictionaries and endangered languages. To appear in D. Bradley and M. Bradley (eds), *Language Endangerment and Language Maintenance: An active approach*. Curzon Press, London [Paper presented at the Applied Linguistics Association of Australia Annual Congress, Perth, 1999, and at the Workshop on Endangered Languages, Melbourne, 1999].

Corris, M., Manning, C., Poetsch, S., and Simpson, J. (2000b). Bilingual Dictionaries for Australian Languages: User studies on the place of paper and electronic dictionaries. *Proceedings of Euralex 2000,* pp. 169–181. Stuttgart.
Corris, M., Manning, C., Poetsch, S., and Simpson, J. (forthcoming). Using dictionaries of Australian Aboriginal Languages. Ms., University of Sydney.
Coward, D.F., and Grimes, C.E. (1995). Making dictionaries: A guide to lexicography and the Multi-Dictionary Formatter (Version 1.0). Summer Institute of Linguistics, Waxhaw, NC.
Eades, P., Huang, M., and Wang, J. (1998). Online Animated Graph Drawing using a Modified Spring Algorithm. *Computer Science '98: Proceedings of the 21st Australian Computer Science Conference*, pp. 17–28. Springer Verlag, Berlin.
GMD-IPSI. (1999). GMD-IPSI XQL Engine version 1.0.2. http://xml.darmstadt.gmd.de/xql.
Goddard, C. and Thieberger, N. (1997). Lexicographic Research on Australian Languages 1968–1993. In M. J. Walsh and D. Tyron (eds), *Boundary Rider: Essays in Honour of Geoffrey O'Grady*, Pacific Linguistics, Canberra, pp. 175–208.
Hale, K. L. (1983). Warlpiri and the grammar of non-configurational languages. *Natural Language & Linguistic Theory* 1: 5–47.
Huang, M. L., Eades, P., and Cohen, R. F. (1998). WebOFDAV: Navigating and visualizing the Web on-line with animated context swapping. *Proceedings of the 7th International World Wide Web Conference*. *Computer Networks and ISDN Systems* 30: 638–642.
Jansz, K., Manning, C. D., and Indurkhya, N. (1999). Kirrkirr: Interactive Visualisation And Multimedia From A Structured Warlpiri Dictionary. *Proceedings of AusWeb99, the Fifth Australian World Wide Web Conference*, pp. 302–316.
Jansz, K., Sng, W. J., Indurkhya, N. and Manning, C. (2000). Using XSL and XQL for Efficient Customised Access to Dictionary Information. *Proceedings of AusWeb2k, the Sixth Australian World Wide Web Conference,* pp. 167–181.
Kegl, J. (1995). Machine-Readable Dictionaries and Education. In D. Walker, A. Zampolli and N. Calzolari (eds) *Automating the Lexicon: Research and Practice in a Multilingual Environment*. Oxford University Press, Clarendon, pp. 249–284.
Laughren, M. and Nash, D. G. (1983). Warlpiri Dictionary Project: Aims, method, organisation and problems of definition. In *Papers in Australian Linguistics No. 15: Australian Aboriginal Lexicography*, Pacific Linguistics, Canberra, pp. 109–133.
Laughren, M., Hale, K., and Hoogenraad, R. (forthcoming). Warlpiri Dictionary. Department of English, University of Queensland, ms.
Miller, G., Beckwith, R., Fellbaum, C., Gross, R., and Miller, K. (1993). Introduction to WordNet: An On-line Lexical Database. In C. Felbaum (ed) *WordNet: An electronic lexical database.* MIT Press, Boston, 1998.
Munzner, T., Guimbretière, F., and Robertson, G. (1999). Constellation: A Visualization Tool for Linguistic Queries from MindNet. *Proceedings of the 1999 IEEE Symposium on Information Visualization,* pp. 132–135.
Pirolli, P., Schank, P., Hearst, M. A., and Diehl, C. (1996). Scatter/Gather Browsing Communicates the Topic Structure of a Very large Text Collection. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '96)*, pp. 213–220.
plumbdesign. (1998). Visual Thesaurus Java applet. http://www.plumbdesign.com/thesaurus.
Schechter, J. B. (1997). The Reader: Interface and Implementation. Ms., Princeton University. http://www.cogsci.princeton.edu/~wn/current/readerdoc.html.
Sharpe, P. (1995). Electronic Dictionaries with Particular Reference to an Electronic Bilingual Dictionary for English-speaking Learners of Japanese. *International Journal of Lexicography,* 8: 39–54.

Simpson, J. H. (1991). Warlpiri Morpho-Syntax: A Lexicalist Approach. Kluwer, Dordrecht.
Sperberg-McQueen, C. M. and Burnard, L. (eds) (1994). Guidelines for electronic text encoding and interchange. Document TEI P3, ms., University of Illinois, Chicago. http://www.uic.edu/orgs/tei/.
Wallace, R., Soloway, E., Krajcik, J., Bos, N., Hoffman, J., Hunter, H. E., Kiskis, D., Klann, E., Peters, G., Richardson, D., and Ronen, O. (1998). ARTEMIS: Learner-Centered Design of an Information Seeking Environment for K-12 Education. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '98),* pp. 195–202.
Weiner, E. (1994). The Lexicographical Workstation and the Scholarly Dictionary. In B.T.S. Atkins and A. Zampolli (eds) *Computational Approaches to the Lexicon*, Oxford University Press, Clarendon, pp. 413–439.
XML. (1998). Extensible Markup Language (XML) 1.0 W3C Recommendation. T. Bray, J. Paoli, and C. M. Sperberg-McQueen (eds). http://www.w3.org/TR/1998/REC-xml-19980210.
XQL. (1999). XQL (XML Query Language). J. Robie (ed). http://metalab.unc.edu/xql/xql-proposal.html
XSLT. (1999). XSL Transformations (XSLT) Version 1.0 W3C Recommendation. J. Clark (ed). http://www.w3.org/TR/1999/REC-xslt-19991116.