Microsoft and Dictionary Makers: Defining Partnerships

Julian Parish

I am happy to take up the invitation to reply to Joseph Esposito's article, *Dictionaries, another Netscape?* (KDN 10, 2002). Esposito's article is fascinating and provocative in many ways, and I share many of his views about the importance of dictionaries and how they will be used in the future. Where we differ is in his interpretation of Microsoft's interest in lexical data and in what this means for established dictionary publishers.

Commitment to lexical data

Microsoft, as rightly identified by Esposito, "views lexical databases as an aspect of strategic technology", and thus perceives such data in the broadest sense: the definitions and translation equivalents which are central to printed reference dictionaries are only part of the lexical data that Microsoft utilises. Information about spelling, pronunciation and grammar is perhaps even more important in many of Microsoft's products. Whether in established businesses like Office or new areas such as Tablet PC or MS Reader, we make extensive use of wordlists – in look-up dictionaries, yes, but also in spellers, handwriting recognizers, and search and speech recognition engines, to give just some examples. Our requirements for lexical data will continue to develop, as we develop new products and add further localized languages to existing products. For Office 2003, for instance, we have added new localized versions for Catalan and Nynorsk.

Furthermore, I share Esposito's view that

Julian Parish is the Business Manager for Dictionary Central at Microsoft's European headquarters in Paris, responsible for coordinating the planning and acquisition of lexical resources across the company. He studied Modern Languages at Oxford, and taught at the Sorbonne.
julianp@microsoft.com

Charles M. Levine's 'The Coming Boom in English Lexicography: Some Thoughts about the World Wide Web (Part One)' appeared in *Kernerman Dictionary News*, Number 9, July 2001 (http://kdictionaries.com/ newsletter/kdn9-1.html).

Joseph J. Esposito's 'Dictionaries, another Netscape?' appeared in *Kernerman Dictionary News*, Number 10, July 2002 (http: //kdictionaries.com/ newsletter/kdn10-1.html).

Levine's 'The Coming Boom in English Lexicography – Reconsidered (Part Two)', appeared in *Kernerman Dictionary News*, Number 11, July 2003 (http: //kdictionaries.com/ newsletter/kdn11-1.html).

Comment from Joseph J. Esposito: "Mr. Parish does not respond to my piece at all. Microsoft's intentions are irrelevant; what is important are its effects, an ineluctable outgrowth of Microsoft's position in the marketplace. Mr. Parish's illustrations all serve to confirm my thesis."

lexical data will increasingly be accessed from within computer applications, whether in machine-readable form or directly by the end-user. During the 1990s many electronic dictionaries were published, often adapted from existing print dictionaries, and marketed as stand-alone consumer products (e.g. Microsoft Bookshelf). That period has passed, as Gilles-Maurice de Schryver notes: "If there is one single feature likely to be applicable to all [electronic dictionaries] of the future, it is that they will stop functioning as stand-alone products." ('Lexicographers' Dreams in the Electronic-Dictionary Age', in *International Journal of Lexicography* 16.2, 2003).

In future, we should expect to see an increasing range of applications which make use of lexical data. Already, these include spellers in products such as Office, Works, Outlook Web, Access and Hotmail, as well as reference dictionaries (both monolingual and bilingual) which may be accessed in the Encarta Reference Library, online on MSN or through the new Research Pane in Office 2003.

Partnership in development

The parallel that has been drawn between the development of lexical data and Netscape is, I feel, misleading: it is technology – whether for Internet browsing or search – and not data itself which is central to Microsoft's business. Developing our own lexical content (across more than 40 languages at that) is simply not a part of the company's core mission. Why should we seek to develop so many dictionaries ourselves when excellent resources already exist, developed over many years by a range of dictionary publishers? Those resources, moreover, are already available, whereas any new dictionary would require several years' work to create *ab initio*.

The example of the *Encarta World English Dictionary* may also be misleading, insofar as the development of this dictionary, whilst jointly funded by Microsoft, was entirely directed by Bloomsbury Publishing in London. In practice, it has been rare indeed for Microsoft to develop dictionary content itself, other than in the specific area of IT vocabulary, for which Microsoft Press publishes a specialized *Computer Dictionary* (http://www.microsoft.com/ mspress/books/5582.asp).

In all other cases, Microsoft has developed its lexical tools with the help of third party specialists in lexical data. These may be publishers with an established background in print-based reference publishing or newer independent software vendors (ISVs) who have built their businesses supplying lexical data to general software houses such as Microsoft.

In an earlier issue of *Kernerman Dictionary News*, Charles Levine questions what opportunity there is for businesses to make any money from developing new spellers: "Since spell checkers are bundled freely, there is no money to be made and no incentive in developing truly better, more intelligent spell-checking software." ('The Coming Boom in English Lexicography: Some Thoughts about the Worldwide Web (I)', in KDN 9, 2001).

That assertion over-simplifies the case of many Microsoft products: whilst the owner of the lexical data used may not be paid directly by the end-user of the software, that data can be monetized through the license fees ISVs charge to companies like Microsoft. And it is in our interest to continue improving the quality of linguistic tools in new product releases.

The parallel I would suggest is not with Internet browsers, but with the use of mapping data in computer software. Over the past decade this market has changed enormously, with paper maps increasingly giving way to electronic applications, first PC-based solutions such as Microsoft's Autoroute Express, then in-car GPS-based navigation solutions. Those cartographers who think of themselves only as book publishers will certainly see their businesses decline. For those companies, however, who see their value in providing high-quality cartographic data in the required (electronic) form, these applications create new business opportunities. Again, those companies may be existing publishers such as Rand McNally or Michelin, or new providers like NavTech.

Continuing our commitment to partnership

As we look to the future, Microsoft sees more, not fewer, opportunities for publishers to provide lexical data to work with our technologies:

First, we will continue to license lexical data for new or existing applications from partners who can offer us high quality resources.

Secondly, we are increasingly creating new opportunities for publishers to develop and market themselves additional products which integrate with our core applications. Examples of these already include the add-on spellchecker files for Office in areas such as law, medicine or economics, which exist today for Dutch, French and Italian. Or again the

Translation Dictionaries technical article (a form of software development kit) for the bilingual dictionaries in Office 2003: this article – which is available free-of-charge – enables publishers of bilingual dictionaries to adapt their existing content and sell it as a module which is fully integrated in Office (see http://www.microsoft.com/downloads/ details.aspx?FamilyId=38934F90-FB06-4ABF-ABA5-94D16BF813BB&displayl ang=en).

Dictionaries and other lexical data remain a strategic investment for Microsoft, but one that we believe is based on partnership, not exclusion, an opportunity for dictionary publishers and not a threat.

Dictionary tools in Microsoft products

Microsoft Office 2003

● The speller, thesaurus and grammar checker are already an established part of the Proofing Tools in Office.

● It is possible to add support in more than 40 available languages.

● The Language Auto-Detect feature in Word automatically recognizes after a few words the language being used and will switch the speller, thesaurus and grammar checker to that language. Alternatively, the language itself can be specified.

● Words that are not included in the standard speller (e.g. specialized terms or company names) can be added to the user's custom dictionary in each individual Office configuration.

● For French, Dutch and Italian, additional spellchecker files covering specialist vocabulary for science, law, medicine, IT and economics can be downloaded and integrated into the existing speller

● Access to a range of research and reference information is offered without leaving the Office application. The dictionaries available include:
  ● *Encarta World English Dictionary*, developed in association with Bloomsbury Publishing, with 100,000 headwords (US and UK versions);
  ● *Encarta French Dictionary*, built specifically for Microsoft by a development team in France, 45,000 headwords;
  ● a German monolingual dictionary, produced by a leading German dictionary publisher, with 57.000 headwords;
  ● bilingual dictionaries for English to and from several languages including French, German, Italian and Spanish.

More to come

● In 2004-2005 Microsoft will be adding localized versions of Office for many languages, with spellchecking support provided in many cases. These new versions will extend its coverage of the languages of the new member states of the European Union and beyond (for languages such as Macedonian and Afrikaans). Discussions are also under way to offer further specialized spellcheckers for other languages.

Encarta Reference Library 2005

● It is possible to consult a dictionary without opening up Office. The same dictionary content is available in the latest versions of Encarta, featuring one-click access to definitions, synonyms and translations (the exact mix varies by language).

● Encarta Reference Library 2005 is available for English (in US and UK editions), French, German, Spanish, and Dutch. For English, French and German, Microsoft offers the same dictionaries as in the Microsoft Office 2003 Research Service; for Spanish, the prestigious dictionary of the Real Academia de la Lengua Espanola is included, while for Dutch the dictionary is provided by Het Spectrum.

Dictionaries for Pocket PC

● The Microsoft dictionaries available in MS Reader format can be downloaded directly to Pocket PC using Active Sync, including a specially shortened version of the *Encarta World English Dictionary* (with concise definitions in English) and bilingual dictionaries for English to and from French, German, Italian and Spanish:
http://www.microsoft.com/reader/downloads/dictionaries.asp.

Online: MSN

● It is possible to access *Encarta World English Dictionary* on the Internet:
http://encarta.msn.com/encnet/features/dictionary/dictionaryhome.aspx.

Computer dictionaries in print

● *Microsoft Computer Dictionary* covers computer and IT terminology in English, with over 10,000 entries:
http://www.microsoft.com/mspress/southpacific/books/book19087.htm.
● The same in German:
http://mspress.microsoft.de/mspress/product.asp?dept%5Fid=2000&sku=3%2D86063%2D896%2D3.