Farr, F. and O'Keeffe , A. (2004) "Introduction". *Teanga* (*Yearbook of the Irish Association for Applied Linguistics*]. 21. 2-5. [special issue on "Corpora, variation and the language classroom"

Introduction

Fiona Farr and Anne O'Keeffe

The occasion of the first international Inter-Varietal Applied Corpus Studies (IVACS) conference at the University of Limerick, June 2002, has given rise to this special issue of Teanga on *Corpora, Varieties and the Language Classroom.* This conference marked the launch of the IVACS research centre, which brings together researchers whose goal is to add to the description of language variation in context and to explore how the outcomes of this empirical research might apply to pedagogy. Currently, this inter-institutional research centre comprises researchers from Mary Immaculate College; the University of Limerick; Queen's University, Belfast; the University of Nottingham, and the University of Technology, Sydney. In our framework for research the term 'variety' is viewed broadly to encompass not just geographical varieties, but also 'varieties' that result from contexts of use (including register and genre, for example, variation across the language of service encounters, casual conversations, institutional interactions and so on depending on the conversational conditions that pertain). This gives a more refined insight into the connection between language in use and conditions of use through the analysis of language corpora.

Our research based on language corpora, in particular spoken language corpora, aims to record and profile language varieties rather than to codify any 'standard' use of language. In line with researchers involved in the International Corpus of English (ICE)-Ireland project, we are very much 'against the standard ideology and belief that there is one and only one correct form of the language, modelled on a single correct written form' (Kallen and Kirk 2001: 60). Instead, we embrace the richness of spoken language in all its guises, and while we describe frequencies, patterns and recurrences, we also allow for the fascinating insights offered by ambiguities, quirks and exceptional language use.

A number of the papers in this special issue are based on research being undertaken within the IVACS research centre. Farr, Murphy and O'Keeffe report on the design and application of the recently completed one-million word Limerick Corpus of Irish English (L-CIE), a spoken corpus which mirrors the design of the five-million word Cambridge and Nottingham Corpus of Discourse in English (CANCODE). McCarthy and Carter draw on CANCODE to investigate multi-word strings. They show that while many such strings have neither syntactic nor semantic integrity, for example *at the*, *what do you*, many display pragmatic integrity, encoding interactive functions such as hedging, vagueness, discourse marking for example, *and that sort of thing, you know, a couple of*. By way of pedagogic implication, they note that clusters contrast with the low frequency of opaque idiomatic expressions and conclude that high-frequency clusters raise issues around the distinction between lexis and grammar, and support a synthetic view of language production and storage, with implications for the understanding of notions such as fluency and idiomaticity.

Binchy uses a corpus of first-year undergraduate Philosophy essays from an Irish third level university collected over two semesters to examine the relationship between lexico-grammar and grade attainment. He focuses in particular on the personal pronoun *I* across grades and semesters thus providing a context-based functional profile as well as a correlation between these functions and the grade received. His work illustrates the importance of the longitudinal corpus study of undergraduate 'developmental' academic discourse.

Hnin Tun details a unique corpus-based study of spoken Burmese particles. These items share many characteristics of discourse markers. However, this study uncovers the limitations of using an 'English-centred' model for the corpus-based study of all languages. Discourse markers have been mostly studied in English, using models which depend on the notion of discourse markers as freestanding morphemes such as words or phrases, whereas the particles described by Hnin Tun are bound morphemes, many of which do not have one-to-one equivalents in English. Therefore the notion of using a 'word' as the basic unit of analysis is of little benefit for the study Burmese. Hnin Tun's study goes beyond traditional syntactic descriptions of Burmese particles to describe their discourse functions which sheds light on their semantic and/or pragmatic properties.

Childs reports on her research into features of ellipsis in the spoken language of two adult literacy learners. She uses the metaphor of 'interference' to refer to how spoken language may hinder adult learners' reading. Childs suggests that it may be helpful for literacy tutors to be aware of differences between spoken and written syntax since spoken situational ellipses draw attention to the different locus of reference (situational or textual) in spoken and written ellipsis, and therefore the different strategy of

interpretation required in reading. She demonstrates that the use and non-use of ellipsis often have communicative functions in the dialogues that reflect linguistic skill rather than sloppiness or incompetence. Her study is a good example of the in-depth findings that even a small corpus can generate within a case study framework.

Chambers and Kelly use a corpus of over one million words of contemporary written standard French, comprising semi-specialized journalistic texts relating to the introduction of the euro. With the aim of illustrating the potential of this type of corpus as a resource for language teachers and learners, they investigate the lexico-grammar associated with the expression of quantitative data. Their study points to the use of such semi-specialised corpora as a complement to core grammars and textbooks as they provide a rich source of context-specific embodiments of actual language use. The authors also address the question of how information such as this can be made available to teachers and learners in an environment which fosters inductive learning and learner autonomy.

In a study based on a sub-corpus of L-CIE, Clancy looks at the little-studied intimate genre of family discourse. Such data poses many challenges to the analyst, not least of all, the difficulty in acquiring it. Clancy looks in detail at the structure of the exchange in multi-party family discourse, which on the surface seems unwieldy and deconstructed. He illustrates how the traditional Sinclair Coulthard Initiation-Response-Feedback model is unsuitable for the analysis of such data. However, his analysis confirms that later models, in particular the work of Hoey, identify a clear exchange structure in multi-party family discourse.

In a paper based on the Hong Kong Corpus of Spoken English (HKCSE), Cheng and Warren compare ways in which the two sets of participants (Hong Kong Chinese and native speakers of English) ask questions. They note for example, that the Hong Kong Chinese have a stronger tendency to ask questions by means of declarative-mood questions. They investigate the intonation used to achieve this indirect question form and compare it with the native speaker data. Their work illustrates the rich potential of using a prosodically transcribed corpus such as the HKCSE. They conclude that through analysing the intonation of the speakers in the corpus, the analyst will better understand discourse intonation in context and how it works across a variety of speakers and discourse types.

Finally, as guest editors, we would like to thank the authors who contributed papers, the reviewers of these papers, the plenary speakers at the first IVACS conference, Ronald Carter, Susan Conrad, Gwyneth Fox and Michael McCarthy, and a special thanks to Jeffrey Kallen for his support and encouragement throughout this endeavour.