James Thomas
Alex Boulton

Input, Process and Product

Developments in Teaching and Language Corpora

Input, Process and Product

muni
PRESS

Input, Process and Product

Developments in Teaching and Language Corpora

Edited by
James Thomas
Alex Boulton

Masaryk University Press • 2012

Thomas, James and Boulton, Alex, eds.: Input, Process and Product. Developments in Teaching and Language Corpora. First edition. Brno : Masaryk University Press, 2012, p. 352. ISBN 978-80-210-5896-5.

ISBN 978-80-210-7636-5 (online : pdf)
ISBN 978-80-210-5896-5 (paperback)

Input, Process and Product
Developments in Teaching and Language Corpora

Edited by James Thomas and Alex Boulton
Designed and typeset by Marek Procházka
Printed in the Czech Republic by Reprocentrum, Blansko
Published by Masaryk University Press • www.muni.cz/press
First edition • 2012 • Brno, Czech Republic

Contents

Section 3: Learner corpora as language description

Section 4: Learner corpora as language input

Preface

Chris Tribble
King's College,
London University

The preface is a strange genre. The Oxford English Dictionary gives Caxton's 1484 preface to the Subtyl Historyes & Fables of Esope as the first recorded instance, but also comments that a preface is:

> The introduction to a literary work, usually stating its subject, purpose, scope, method, etc.; (in modern use also) *spec.* an introductory note, often of a personal nature, written by the author and distinguished from a foreword and an introduction.

As I'm not the author of this volume, and as Alex Boulton and James Thomas have already written an excellent, and extensive, introduction, I felt at a bit of a loss when starting out to write this present preface. However, in the best tradition of empirical language studies I decided to seek guidance on how I might go about writing into this genre by consulting a range of permissible exemplars. After some searching, I decided to focus on one of the best known English variants: Wordsworth's preface to The Lyrical Ballads – a 9,000-word argument for the kinds of poetics the author wished to present to the public. Clearly, as a corpus linguist of a kind, my next step was to see what corpus analysis might offer to help me in my task. With WordSmith Tools (Scott 2008) to hand, I quickly generated a wordlist for this text, and then a set of keywords (referenced against the British National Corpus). And what did I find apart from the words *poem* and *poetry*? At the top of the list came: *pleasure*, *language*, *reader*, and *passions*. And there was my framework for my preface to Input, Process and Product: Developments in Teaching and Language Corpora.

First, *pleasure*. The TaLC conferences represent the only series of academic gatherings that I've been so consistently engaged with across my academic career. I wasn't at the first one in Lancaster, and, much to my regret, I wasn't able to make TaLC5 in Bertinoro. However, I've been there for all the others, and each one has given me the pleasure of developing friendships which have extended beyond the three days of the conference, and also the pleasure of witnessing the emergence of a community of practice. Through the process of preparing research papers to present at the conference, learning from leading practitioners, and sharing experience, the TaLC series has ensured that I have a regular update on the state-of-the-art in my field. This pleasure has been enhanced by the experience of visiting different countries and their leading universities. On each of these occasions, there have been new and renewed encounters with others who have a shared commitment to discovering how best to use texts and computer tools to meet the needs of a widening population of learners. At TaLC9 in Brno, some excellent Czech beers and wines added a further pleasure to the process.

In terms of *language*, TaLC9 also lived up to all my expectations. As you will see from the papers in this collection, work in our field now has a much greater emphasis on classroom realities, and on the use of learner language as the starting point for investigation, than was the case ten or fifteen years ago. As language data has become more readily available (in the form of standardised, publically available corpora, the smörgåsbord of the Internet, or small, tailored specialist corpora), computers have become more powerful, and as off-line and on-line corpus tools have become easier to find and easier to use, researchers, teachers, and materials authors have started to provide practical responses to learner needs. John Sinclair's earlier vision (expounded at TaLC1 in 1994) where "quite young learners will gain access to this and will become self-taught DDL

(data-driven learning) students" (Sinclair 1997: 30) may not have been fulfilled, but this collection does give clear accounts of how students in classrooms as far apart as Michigan, Portugal and Shanghai are being supported through the use of language corpora.

From the *reader's* perspective, this selection of papers from the 9th conference in the TaLC series constitutes an essential update to the baseline of where we are in our professional development. I use the word 'development' advisedly, as although I am not a great believer in modernist views of progress, I do hold that the spectacular technological changes which we have experienced in the sixteen years between TaLC1 and TaLC9 have had a profound impact on our ability to exploit language data to pedagogic ends. The different volumes which have reported on the conference series give the reader a clear account of how access to corpus data has been opened to unimaginably wider communities, and how new groups of students are able to benefit from this access. The process is not yet complete (how can anything associated with language ever be considered complete?), but the reported experience given to the reader through the series of TaLC conference publications provides an essential account of what has happened (see the appendix to the introduction to this volume by Boulton and Thomas). Whatever we do in our next research project or lesson plan will be enhanced by our awareness of our own professional history. In this respect, this volume is an essential part of a larger story.

And finally, *passions*. I have to admit that *passion* is a word which is now, for me, tainted by its overuse in inspirational management-speak and media discourses. The example in Figure 1

Figure 1. *Passion* in UK newspapers

Going to drama college reconfirmed my	passion	for acting and then this stage w
diately connected as we both shared a	passion	for social issues. We started co
It may be flawed, but there's genuine	passion	at the heart of The Iron Lady, w
a gallop through a life forged by her	passion	for politics — starting in her t
ou roll up your sleeves and show some	passion.	When you are against the wall y
nspirational, with the enthusiasm and	passion	for open access from African res
az Manzoor had the floor. He spoke of	passion	and inspiration, of the courage
eaches us to resolve to lead lives of	passion	and conviction," he said. It loo
ey don't care for Dinamo. No fire, no	passion.	I remember once when I was at s
does tend to rather get in the way of	passion.	Strange's speech to Lund at her
acerbic critic and broadcaster with a	passion	for literature and art, he is kn
about this club and very obviously a	passion.	I would consider it a major par
major part of my job to reignite that	passion.	It's so exciting, I can hardly
y boyhood team but there's incredible	passion	around the place. They turn up i
tinez/Reuters Rafael Nadal ran out of	passion,	and hit the road for Mallorca.
he admitted he had a "little bit less	passion	for the game". But Djokovic, too
raud can occur] so you need to have a	passion	and to know it can be successful
belt context. Acknowledging that his	passion	for the "sandcastle" qualities o
in statistics and probabilities, his	passion	for betting on the beautiful gam

was generated from UK news sources using WebCorp (http://www.webcorp.org.uk) and gives a flavour of its current usage.

However, there is something peculiarly passionate about the way in which anyone who gets involved in corpus-informed language teaching will put in hours that are well outside their normal job description in order to produce teaching materials which they know are grounded in observed reality. To know that what you are doing is built on language data which is open to challenge, available to others to test and to learn from, and which provides students with an account of language which is wider than the sum of your own experience (and prejudices) is intensely satisfying. This passion is present in all of the papers in this collection, and it will, I am convinced, continue to help us to carry forward our work until the next TaLC conference and, I trust, the next collection of selected papers.

London, December 2011

References

Scott, M. 2008. *WordSmith Tools* 5.0. Liverpool: Lexical Analysis Software.

Sinclair, J. M. 1997. Corpus evidence in language description. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (eds), *Teaching and Language Corpora*. London: Longman, p. 27-39.

Introduction: Corpus language input, corpus processes in learning, learner corpus product

Alex Boulton
Crapel – ATILF / CNRS,
University of Lorraine

James Thomas
Department
of English and
American Studies,
Faculty of Arts,
Masaryk University

Corpus linguistics is essentially concerned with describing language for linguistic research purposes, but language corpora (along with the associated tools and methodologies) have many different affordances and applications. In the field of language teaching, corpus analysis is used to inform the content decisions of what to teach different learner populations in different contexts at different stages of development. This typically includes the application of frequency data in determining the sequence in which linguistic items should be introduced, in identifying key multi-word units and a wide range of lexico-semantic patterns, and in predicting areas of potential difficulty from learner corpora. This essentially indirect approach (Römer 2011) to corpus data is taken by syllabus designers, materials writers, lexicographers and testers, though the results may be entirely invisible to the end user (McCarthy 2004). However, teachers can also make use of corpora to answer their own questions about language, to test grammar 'rules' against real data, to find examples and help create materials for teaching and testing, among other things. Learner involvement need not be limited to teacher mediated uses, but can involve direct hands-on consultation, either for language learning or as a reference resource. This is commonly associated with the work of Tim Johns[1] in what he called data-driven learning (DDL), an approach which conflates the roles of learners and researchers and sees them deriving their own answers from direct contact with the data (e.g. Johns & King 1991). The approach is essentially constructivist, providing an authentic way of tackling lexico-grammar in particular (Thomas 2006) in contrast to most decontextualised and relatively 'artificial' vocabulary learning techniques – assuming any strategies are taught at all.

The papers selected for inclusion in this volume derive from presentations given at TaLC9 in Brno in 2010, 16 years after the first TaLC conference was held in Lancaster in 1994. Looking through the list of over 150 papers published from almost two decades of TaLC conferences (see Appendix), an evolutionary trajectory emerges: while many of the early issues are still relevant today, other have opened up in various ways, and this volume includes some papers that cover entirely new ground. TaLC is thus no longer in its infancy – but neither has it reached full maturity. It has gone beyond the initial idea of concordancing by advanced adult L2 students for lexico-grammar (e.g. Tribble & Jones 1997), to being employed in an ever-expanding array of linguistic fields from discourse analysis (e.g. Charles 2007) to literary studies (e.g. Kettemann & Marko 2004, 2011) to translation (e.g. Kübler 2003), at lower levels (e.g. Cobb et al. 2001), in schools (e.g. Sun & Wang 2003) and even for primary schools for L1 (e.g. Sealey & Thompson 2007). The early research enthusiasm is as strong as ever and is constantly passed on to generations of new researchers, but given the development of new types of corpora, of more sophisticated software and of computer technology in general, there can be no certainties about what directions it is likely to take, nor how it may eventually earn its keep in regular classroom practice. Despite the considerable technological advances and numerous publications in the spheres of language education, it is frequently remarked that TaLC remains marginal to mainstream language teaching (e.g. Chambers et al. 2011).

One probable cause of this lack of uptake in mainstream language education is that TaLC, at least in popular perception, remains stubbornly the province of researchers rather than teachers,

1 1936-2009. See the obituary by Scott (2009).

let alone learners (Mukherjee 2004), a gap that desperately needs to be bridged (cf. McCarthy 2008). Worse, corpus work is seen as an ivory tower activity, generating a notable lack of empirical classroom research (e.g. Johansson 2009; Yoon 2011). However, a growing body of studies do attempt to evaluate some aspect of corpus use in real classroom contexts – 93 separate studies to date, according to a current survey by Boulton (2010). These are tremendously varied in design, underlining the flexibility of approaches to corpus use for a variety of different learner needs in very different conditions; as Breyer (2006: 162) has pointed out, corpus activities are "limited only by the imagination of the user." But regardless of how corpora are introduced, the overwhelming conclusion is that learners can use them effectively for many different purposes, are receptive to the approach and see the relevance to their own needs, and can use them successfully both as a learning tool and as a reference resource, particularly for writing, revision, error-correction and translation.

The TaLC conference series combines, as its name suggests, teaching and language corpora. But crucially, teaching is the first of the two terms, and this is reflected in the structure of the present volume, with the first two sections looking at how corpora can be used as *input* for language learning. Section One opens with a paper by Ana Frankenberg-Garcia, who asks why corpus use is not more widespread among the language teaching community, and provides a number of suggestions for how corpora can be integrated into everyday language classes. For her, the crucial issue is not what teachers and learners can do with corpora, but what corpora can do for teachers and learners. The remaining chapters in this section explore some of the potential for corpus use in language teaching. Patrick Hanks combines prototype theory and corpus linguistics to show how pattern analysis can lead to a radically different approach to language and linguistics, in the process transforming dictionaries and other reference resources for language teachers and learners. The result is firmly rooted in actual language use, integrating focus on form and on meaning into a fundamentally innovative tool for these end users. Teachers and learners can also exploit corpora as a reference resource, as discussed largely in Section 2, but a number of initial considerations in developing corpora and software are reported in the next two papers in this section. Shozo Yokoyama, Chizuko Suzuki, Seisuke Yasunami and Naoko Kawakita describe the construction of a corpus of academic research articles in medicine, which they analyse for different types of verbs. It is argued that learners can benefit from the resulting insights in terms of frequency, keyness, collocates and distributions over different IMRAD sections, which they can discover using the dedicated corpus interface outlined in the paper. Ute Römer also describes a specialised corpus and interface, but here compiled from high-scoring essays mainly by native speakers who are still learning their own discipline. The Michigan Corpus of Upper-level Student Papers (MICUSP) is thus pedagogically relevant to EAP learner / apprentice writers: teachers can use it to inform their teaching, and learners can explore it in a DDL approach to academic writing through a simple on-line interface, as the paper reports. MICUSP is the written counterpart to MICASE (the Michigan Corpus of Academic Spoken English), and though corpora of spoken language are more difficult to compile than those of written language, they are of great importance in developing the teaching of oral skills. To this end, Stefanie Dose shows that a corpus of TV transcripts can be tremendously valuable for pedagogical purposes, demonstrating that the language is in many ways remarkably similar to unscripted speech. TV series can provide a corpus that learners can relate to or 'authenticate' (cf. Widdowson 2000), and allow work on individual

written or multimedia extracts for a variety of activities – a "pedagogically relevant" corpus in Braun's (2005) terms. While we can certainly subvert linguistic corpora for language teaching, this inevitably involves a certain amount of "rethinking" (Burnard & McEnery 2000).

These introductory chapters derive from the contributors' many years of experience in using corpus data either directly or indirectly for language learning – they are far from ivory tower expositions divorced from reality. Section Two makes the connection between corpus and classroom more explicit: all of the contributions report on actual applications and evaluate outcomes, attitudes and behaviours of learners faced with corpora and associated tools – the processes involved in using corpora in language teaching and learning.

A recurring question is how corpus work can be successfully integrated into normal classroom practice, as highlighted in the paper by Monika Geist[2] and Angela Hahn. Their results are encouraging insofar as their learners are clearly able to use the general British National Corpus (BNC) for specific ends with some success, even though some of them lacked the necessary motivation to invest time and effort in corpus activities which were not graded and which the learners were unable to relate to their regular classes. It is common practice to introduce corpus activities as an add-on, going against the precept of constructive alignment (e.g. Biggs 1996). But DDL can be introduced as 'ordinary' practice as demonstrated in the study by Henry Tyne, who shows that it is perfectly compatible with standard teaching techniques – including at the level of text. The teachers in his study report that the DDL techniques involved are of immediate benefit in their daily teaching, and may even provide a way in to more usual DDL activities later on. Another option is for the teacher to mediate the corpus data and use only printed materials, thus eliminating the 'obstacle' of the computer in DDL. Alex Boulton reports on using DDL with and without a computer, finding that each approach has its own advantages in terms of learning outcomes and appeals to different learners. In a similar vein, Kiyomi Chujo and Kathryn Oghigian find that optimal results may be obtained from a combination of paper-based and computer-based DDL, here in terms of feedback and learning outcomes for vocabulary and grammar. Examples such as these show that corpora can be easily and efficiently exploited by learners even without extensive training in the associated tools. This is confirmed in the following paper by Klára Osolsobě and Pavlína Vališová, where learners of Czech managed to conduct simple queries and obtain meaningful results with a minimum of training. Even the seemingly complex work with lexical bundles reported by Andreas Eriksson was conducted over only two workshop sessions, suggesting that focusing on specific tasks in relevant specialist fields can make corpus work more relevant and motivating and thus more accessible.

These first two sections show that corpus use is no longer the sole preserve of the "particular type of student" typical of early DDL work – "adult: well-motivated, a sophisticated learner with experience of research methods in his subject area with particular needs... in a particular learning/teaching situation" (Johns 1986: 161). This evolution is perhaps inevitable with the increasing availability of a variety of corpora and more user-friendly software, appropriate even for secondary school students as exemplified in the studies by Geist and Hahn as well as by Tyne (where the teachers are also regular teachers and not researchers). Though it is true that many of the studies

2 Monika Geist originally contributed to this paper as Monika Formánková.

here do involve undergraduates, most are students who are not majoring in languages, often with low levels of motivation, little sophistication in language learning, and relatively low levels of proficiency – pre-intermediate in Boulton, beginners in Chujo and Oghigian.

While English is perhaps inevitably the most common target language, Tyne's students are learning Spanish, Osolsobě and Vališová's learning Czech (one cohort even consists of native speakers), underscoring the flexibility of corpus-based activities even for languages which are quite different from English in terms of morphological complexity and syntax. The types of data used also vary widely, from four million words of general English in Geist and Hahn to the level of individual text in Tyne; from student papers in Römer to expert writing in Eriksson and Yokoyama et al.; parallel corpora in Chujo and Oghigian; spoken data in Dose, and so on. The tasks and types of analysis are correspondingly varied, from the very simple lexical level for younger learners in Geist and Hahn to lexical bundles in Eriksson and phraseology in Römer. The overall picture which emerges is that corpora and DDL hold something for everyone: there is no 'best' corpus for all purposes and no exclusive 'right' way to exploit corpora: pedagogical relevance and appropriateness in each specific case is paramount (Flowerdew 2009).

Sections 3 and 4 move on to learner corpora, i.e. corpora compiled from the spoken or written *output* of learners, which can be quantified and analysed in the same way as corpora consisting of native or expert texts (Leńko-Szymańska 2008). The results serve many purposes as can be seen from the wide variety of issues covered here, reflecting the burgeoning field of learner corpus research spanning the last 20 years (cf. Granger 2009). As with corpora of native speaker or expert texts, learner corpora can be used in a data-driven learning approach (Granger & Tribble 2006) where learners analyse corpora comprising texts of their own language output or those of others (Seidlhofer 2000). They are also valuable in the automatic detection of errors and the automatic correction and scoring of student writing. They can be used to inform materials, resources and practices as well as testing and assessment tools. They can improve our knowledge of the processes involved in language acquisition and interlanguage development, and allow us to relate particular features to different levels of proficiency. In the classroom, they are a resource for systematically raising teachers' awareness of their own learners' specific problems, while also exemplifying the successful use of the features of student output that can be observed and used as models of good practice.

But probably the most frequent approach, and the one that launches Section Three, is the comparison of learner and native corpora, usually with a focus on 'errors' – including the under- and overuse of various linguistic features. Corpus linguistics allows rigorous analysis of learner output for systematic detection and exploration of areas of difficulty where previous attempts could rely on little more than a hunch based on personal experience or intuition; it is therefore unsurprising that contrastive analysis has made something of a comeback in recent years. Several papers here thus attribute different error types directly to the learner's mother tongue (L1), potentially an argument for a return to the use of materials produced with the specific L1 in mind and against the use of generic textbooks produced by international publishers for global distribution.

Marina Mattheoudakis and Anna-Maria Hatzitheodorou compare learner writing against native texts for collocates of delexical or 'light' verbs. Their analysis suggests that transparency and the existence of comparable collocates in the L1 are major factors in predicting erroneous as

well as over- and underused collocates; without them, learners have little choice but to rely on Sinclair's (1991: 109ff) "open-choice principle" rather than his "idiom principle". As such items tend to lack salience, training is needed in noticing. This is the case for many spoken features too, as shown in the paper by Sandra Götz who finds that even advanced learners tend to speak less (in terms of words per minute or length of turn) than native speakers, and exhibit greater use of unfilled pauses and other hesitation phenomena along with more limited use of discourse markers. A final paper comparing learner and native corpora also looks at discourse markers in speech: Jiajia Xu, Mark Morgan and John McKenny highlight the need for intuition in complementing automatic extraction of semantically relevant n-grams. Differences are again attributed largely to L1 transfer, with overuse in particular being linked to a more limited repertoire of connectors due in part to decontextualised overteaching of specific items. A similar point is made by Svetla Rogatcheva, who contrasts required and optional contexts for different verb aspects in the present and past, showing that Bulgarian learners have more difficulty with the English progressive, German learners with the perfect. These problems can be linked not only to the L1, but also again to overteaching which might deter learners from using items perceived as problematic. Most of these papers are based on existing learner corpora, but Sylwia Twardo shows that it is possible to create even a fairly large (300,000-word) PoS-tagged learner corpus from scratch. She takes up a theme mentioned by Rogatcheva and Xu et al., namely the difficulties involved in dealing with automatic error-detection. These are most visible in the form of 'non-words' arising from spelling or morpheme errors, which occur fairly predictably across different levels of proficiency.

Such contrastive analyses are certainly useful, but the authors do not claim that every difference between native and non-native use is an error to be eradicated at the earliest opportunity: there is often a good reason underlying interlanguage differences (Aston 2008). For example, the presence or overuse of some features (e.g. full forms instead of contractions, overuse of connectors or temporal markers) may increase communicative effectiveness if they in fact compensate for other difficulties (e.g. mastery of pronunciation, deixis or tenses respectively). Similarly, the absence or underuse of particular items (e.g. complex sentence structures or phrasal verbs) may also be communicatively more effective at early stages of development (cf. Larsen-Freeman & Cameron 2008). Finally, learners may even be more effective than monolingual native speakers in intercultural contexts where they may, for example, use fewer idioms or opaque expressions, and be more direct in speech acts such as disagreeing or asking for help (cf. Barbour 2004). While it is important to note such differences, for all these reasons care should be taken to distinguish features that significantly impede communication, those that have little if any effect, and those that may actually be advantageous (cf. Seidlhofer 2011). The point being made here is that the value of learner corpora goes beyond mere error analysis, and it is as important to see what learners *can* do as what they can't – all, of course, for different learners in different conditions at different stages of development (cf. the earlier discussion of MICUSP by Römer).

These are some of the issues taken up in the final section of highly innovative papers, beginning with the article by Susanne Kämmerer: although she also discusses errors in a series of studies, this is crucially from the learner's perspective. Three years after the compilation of the corpus, the original German contributors were able to detect their own errors in only 30% of cases; however, they were able to correct almost all errors once they were pointed out and to explain most, attributing them overwhelmingly to L1 interference or 'stupid mistakes'. Such

insights are important, as the inevitable question is what a teacher should do with errors once they have been detected. M. Trevor Shanklin addresses this issue in considering how automatically generated feedback from oral exams should be useful not just to test-designers and examiners but also to test-takers. This is the aim of the corpus in the Computer Assisted Screening Tool (CAST): basic information such as type/token ratio and mean length of utterance are discussed in relation to proficiency, as are more specific features such as the appropriate use of tenses and subordination. While much of this still focuses on errors, the intention is for the corpus to further serve as an indicator of what successful learners can actually do at different levels, an assumption underpinning the English Vocabulary Profile lists analysed in the final paper by Yukio Tono. The underlying idea of the English Profile project (now with its own journal) is to provide detailed descriptions of what learners of English show they can do at different levels rather than identifying what they get wrong (i.e. what they *should* know). This laudable aim is inevitably fraught with difficulties, as Tono's analysis reveals: in particular, the procedures for deriving the lists from the very large Cambridge native and learner (exam) corpora are not entirely transparent, and it is difficult to attribute different levels to the different senses and uses of individual items. The problems are similar in this respect to the sequencing of dictionary entries, but it is argued that particular attention needs to be paid to receptive and productive uses.

Most of the papers in these sections on learner corpora use a published corpus, especially one of those made available at the Centre for English Corpus Linguistics (CECL) at the Université Catholique de Louvain[3], namely the International Corpus of Learner English (ICLE) and the Louvain International Database of Spoken English Interlanguage (LINDSEI). The former consists of written texts in the form of argumentative essays, the second of traditional oral exam-style questions. One advantage of this suite of corpora is that it is possible to focus on a sub-corpus of learners according to their L1: CECL sub-corpora from Bulgarian, French, German, Greek and Spanish learners all feature in the papers here, along with L1 Chinese and Polish from other sources. Only Shanklin and Tono use learner corpora from speakers of different L1s, but for very explicit reasons: in the former, to produce tools that can be used for different target languages; in the latter to explore a generic, non-language specific resource from a major publisher.

ICLE and LINDSEI can each be compared against an equivalent native-speaker corpus also produced by the CECL: the Louvain Corpus of Native Speaker English Essays (LOCNESS), and the Louvain Corpus of Native Speaker Conversation (LOCNEC) respectively – the former used in Mattheoudakis and Hatzitheodorou, the latter in Götz. The learner corpora are undoubtedly 'authentic' even though the data are gathered in highly controlled conditions, as the contexts reflect 'typical' learner communicative contexts – participating in written and oral exams (cf. Mendikoetzea et al. 2010: 183). While the native speaker corpora might be considered less authentic (or at least, less ecological, as native speakers do not necessarily participate in similar types of exams), it clearly makes sense to compare learner language against native language gathered in comparable situations. However, other corpora such as MICASE or the BNC are for many purposes sufficiently comparable (as here in Xu et al.).

3 See http://www.uclouvain.be/en-cecl.html, accessed 20/11/11.

TaLC, then, is maturing nicely. Kudos must of course go to the visionary pilgrim fathers who made the connection between esoteric linguistic research and the overwhelmingly practical concerns of language teaching and learning, but the ever-expanding CV of TaLC-related publications[4] bears testament to growing research interest around the world. And not just research: the various corpora at Brigham Young University are accessed by over 80,000 individual users each month; of these, only 15% declare their main interest in corpora as being for research purposes (in linguistics, sociology, cultural studies, literature and politics); 28% for professional uses (translators, writers, lexicographers and testers). 15% are teachers (native and non-native), but the largest group by far consists of language learners at 42%.[5] This augurs well for further developments relating teaching and language corpora, an area to which this volume makes its own contribution.

The present volume would not have been possible without the input of certain individuals and organisations. First among these is the TaLC organising committee who blind-reviewed the papers prior to the Brno conference (2010) as well as all full submissions to this volume: Guy Aston, Lou Burnard, Lynn Flowerdew, Bernhard Kettemann, Natalie Kübler, Agnieszka Leńko-Szymańska, Ute Römer and Christopher Tribble. We are also enormously grateful to Marek Procházka, a doctoral student in the Faculty of Arts at Masaryk University, for his typesetting of the whole book.

References

Aston, G. 2008. It's only human... In A. Martelli & V. Pulcini (eds), *Investigating English with Corpora: Studies in honour of Maria Teresa Prat*. Monza: Polimetrica, p. 343-354.

Barbour, S. 2004. Do English-speakers really need other languages? In K. Malmkjaer (ed.), *Translation in Undergraduate Degree Programmes*. Amsterdam: John Benjamins, p. 185-195.

Biggs, J. 1996. Enhancing teaching through constructive alignment. *Higher Education* 32: 347-364.

Boulton, A. 2010. Learning outcomes from corpus consultation. In M. Moreno Jaén, F. Serrano Valverde & M. Calzada Pérez (eds), *Exploring New Paths in Language Pedagogy: Lexis and corpus-based language teaching*. London: Equinox, p. 129-144. Updated supplement (description of 93 empirical DDL studies) at CorpusCALL: http://corpuscall.eu/course/view.php?id=5#sectionblock-2, accessed 15/09/11.

Braun, S. 2005. From pedagogically relevant corpora to authentic language learning contents. *ReCALL* 17(1): 47-64.

Breyer, Y. 2006. My Concordancer: Tailor-made software for language learners and teachers. In S. Braun, K. Kohn & J. Mukherjee (eds), *Corpus Technology and Language Pedagogy: New resources, new tools, new methods*. Frankfurt: Peter Lang, p. 157-176.

Burnard, L. & T. McEnery (eds). 2000. *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang.

4 See the bibliographical database at CorpusCALL, the corpus-related Eurocall SIG: http://corpuscall.eu/, accessed 24/11/11.

5 http://corpus.byu.edu/, accessed 24/11/11. Figures kindly provided by Mark Davies, personal communication.

Chambers, A., F. Farr & S. O'Riordan. 2011. Language teachers with corpora in mind: From starting steps to walking tall. *Language Learning Journal* 39(1): 85-104.

Charles, M. 2007. Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes* 6(4): 289-302.

Cobb, T., C. Greaves & M. Horst. 2001. Peut-on augmenter le rythme d'acquisition lexicale par la lecture? Une expérience de lecture en français appuyée sur une série de ressources en ligne. In P. Raymond & C. Cornaire (eds), *Regards sur la Didactique des Langues Secondes*. Montréal: Editions Logique, p. 133-153. [Translation: Can the rate of lexical acquisition from reading be increased? An experiment in reading French with a suite of on-line resources.] http://www.lextutor.ca/cv/, both accessed 01/12/11.

English Profile Journal. [on line] http://journals.cambridge.org/action/displayJournal?jid=EPJ, accessed 24/11/11.

Flowerdew, L. 2009. Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics* 14(3): 393-417.

Granger, S. 2009. The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (ed.), *Corpora and Language Teaching*. Amsterdam: John Benjamins, p. 13-32.

Granger, S. & C. Tribble. 1998. Learner corpus data in the foreign language classroom: Form-focused instruction and data-driven learning. In S. Granger (ed.), *Learner English on Computer*. London: Longman, p. 199-209.

Johansson, S. 2009. Some thoughts on corpora and second-language acquisition. In K. Aijmer (ed.), *Corpora and Language Teaching*. Amsterdam: John Benjamins, p. 33-44.

Johns, T. 1986. Micro-Concord: A language learner's research tool. *System* 14(2): 151-162.

Johns, T. & P. King (eds). 1991. *Classroom Concordancing*. *English Language Research Journal* 4.

Kettemann, B. & G. Marko. 2004. Can the L in TaLC stand for literature? In G. Aston, S. Bernardini & D. Stewart (eds), *Corpora and Language Learners*. Amsterdam: John Benjamins, p. 169-193.

Kettemann, B. & G. Marko. 2011. Data-driving critical discourse analysis. In N. Kübler (ed.), *Corpora, Language, Teaching, and Resources: From theory to practice*. Bern: Peter Lang, p. 19-48.

Kübler, N. 2003. Corpora and LSP translation. In F. Zanettin, S. Bernardini & D. Stewart (eds), *Corpora in Translator Education*. Manchester: St Jerome Publishing, p. 25-42.

Larsen-Freeman, D. & L. Cameron. 2008. *Complex Systems and Applied Linguistics*. Oxford: Oxford University Press.

Leńko-Szymańska, A. 2008. Non-native or non-expert? The use of connectors in native and foreign language learners' texts. *Acquisition et Interaction en Langue Etrangère* 27: 99-108.

McCarthy, M. 2004. *Touchstone: From corpus to coursebook*. Cambridge: Cambridge University Press. http://www.cambridge.org/other_files/downloads/esl/booklets/McCarthy-Touchstone-Corpus.pdf, accessed 07/10/11.

McCarthy, M. 2008. Accessing and interpreting corpus information in the teacher education context. *Language Teaching* 41(4): 563-574.

Mendikoetzea, A., S. Bielsa & P. Rollinson. 2010. Focus on errors: Learner corpora as pedagogical tools. In M-C. Campoy, B. Bellés-Fortuño & M-L. Gea-Valor (eds), *Corpus-Based Approaches to English Language Teaching*. London: Continuum, p. 180-194.

Mukherjee, J. 2004. Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In U. Connor & T. Upton (eds), *Applied Corpus Linguistics: A multidimensional perspective*. Amsterdam: Rodopi, p. 239-250.

Römer, U. 2011. Corpus research applications in second language teaching. *Annual Review of Applied Linguistics* 31: 205-225.

Scott, M. 2009. In memory of Tim Johns. *International Journal of Corpus Linguistics* 14(3): 271-274.

Sealey, A. & P. Thompson. 2007. Corpus, concordance, classification: Young learners in the L1 classroom. *Language Awareness* 16(3): 208-223.

Seidlhofer, B. 2000. Operationalizing intertextuality: Using learner corpora for learning. In L. Burnard & T. McEnery (eds), *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang, p. 207-223.

Seidlhofer, B. 2011. *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sun, Y-C. & L-Y. Wang. 2003. Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning* 16(1): 83-94.

Thomas, J. 2006. Using corpora in language teaching and learning. *Teaching English with Technology* 6(1): 44-56. http://bit.ly/TEWT2006_Thomas, accessed 24/10/11.

Tribble, C. & G. Jones. 1997. *Concordances in the Classroom*. 2nd edition. Houston: Athelstan.

Widdowson, H. 2000. On the limitations of linguistics applied. *Applied Linguistics* 21(1): 3-25.

Yoon, C. 2011. Concordancing in L2 writing class: An overview of research and issues. *Journal of English for Academic Purposes* 10: 130-139.