



Multiple Affordances of Language Corpora for Data-driven Learning

EDITED BY

Agnieszka Leńko-Szymańska

Alex Boulton

Studies in Corpus Linguistics

69

JOHN BENJAMINS PUBLISHING COMPANY

Multiple Affordances of Language Corpora for Data-driven Learning

Edited by

Agnieszka Leńko-Szymańska

University of Warsaw

Alex Boulton

ATILF-CNRS / University of Lorraine

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Cover design: Françoise Berserik

Cover illustration from original painting *Random Order*

by Lorenzo Pezzatini, Florence, 1996.

DOI 10.1075/scl.69

Cataloging-in-Publication Data available from Library of Congress:

LCCN 2014049652 (PRINT) / 2015006473 (E-BOOK)

ISBN 978 90 272 0377 9 (HB)

ISBN 978 90 272 6871 6 (E-BOOK)

© 2015 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands

John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Multiple Affordances of Language Corpora for Data-driven Learning

Studies in Corpus Linguistics (SCL)

ISSN 1388-0373

SCL focuses on the use of corpora throughout language study, the development of a quantitative approach to linguistics, the design and use of new tools for processing language texts, and the theoretical implications of a data-rich discipline.

For an overview of all books published in this series, please see
<http://benjamins.com/catalog/books/scl>

General Editor

Elena Tognini-Bonelli
The Tuscan Word Centre/
The University of Siena

Consulting Editor

Wolfgang Teubert
University of Birmingham

Advisory Board

Michael Barlow
University of Auckland

Douglas Biber
Northern Arizona University

Marina Bondi
University of Modena and Reggio Emilia

Christopher S. Butler
University of Wales, Swansea

Sylviane Granger
University of Louvain

M.A.K. Halliday
University of Sydney

Yang Huizhong
Jiao Tong University, Shanghai

Susan Hunston
University of Birmingham

Graeme Kennedy
Victoria University of Wellington

Michaela Mahlberg
University of Nottingham

Anna Mauranen
University of Helsinki

Ute Römer
Georgia State University

Jan Svartvik
University of Lund

John M. Swales
University of Michigan

Martin Warren
The Hong Kong Polytechnic University

Volume 69

Multiple Affordances of Language Corpora for Data-driven Learning
Edited by Agnieszka Leńko-Szymańska and Alex Boulton

Editors' acknowledgements

Inevitably, a book such as this does not happen by itself. The editors wish to thank the authors of all the chapters who have patiently responded to our comments, as well as the contributors of other chapters which do not in the end appear in this volume – though the quality in many cases was tremendous, difficult choices had to be made based on the specific DDL focus of this book. All proposals were submitted to a double blind review, and we would like to extend special thanks to all the reviewers from 14 different countries in Europe and around the world: Virginie André, Laurence Anthony, Guy Aston, Angela Chambers, María-Belen Diez-Bedmar, Fiona Farr, John Flowerdew, Lynne Flowerdew, Ana Frankenberg-Garcia, Gaëtanelle Gilquin, Sandra Götz, Nick Groom, Bernhard Kettemann, Natalie Kübler, David Lee, Fanny Meunier, María Moreno Jaén, Pascual Pérez-Paredes, Randi Reppen, Ute Römer, Mike Scott, Simon Smith, Paul Thompson, Yukio Tono, Chris Tribble and Henry Tyne. We would also like to thank the excellent people at John Benjamins for their confidence in us, and all those who have contributed to make TaLC conferences the friendly and stimulating events they are.

Agnieszka Leńko-Szymańska & Alex Boulton

Table of contents

Editors' acknowledgements	VII
Introduction: Data-driven learning in language pedagogy <i>Agnieszka Leńko-Szymańska & Alex Boulton</i>	1
Data-driven learning and language learning theories: Whither the twain shall meet <i>Lynne Flowerdew</i>	15
Teaching and language corpora: Perspectives from a personal journey <i>Christopher Tribble</i>	37
PART I. Corpora for language learning	
Learning phraseology from speech corpora <i>Guy Aston</i>	65
Stealing a march on collocation: Deriving extended collocations from full text for student analysis and synthesis <i>James Thomas</i>	85
A corpus and grammatical browsing system for remedial EFL learners <i>Kiyomi Chujo, Kathryn Oghigian & Shiro Akasegawa</i>	109
PART II. Corpora for skills development	
Same task, different corpus: The role of personal corpora in EAP classes <i>Maggie Charles</i>	131
Textual cohesion patterns for developing reading skills: A corpus-based multilingual learning environment <i>Svitlana Babych</i>	155
Exploiting keywords in a DDL approach to the comprehension of news texts by lower-level students <i>Alejandro Curado Fuentes</i>	177

PART III. Corpora for translation training

Webquests in translator training: Introducing corpus-based tasks <i>Teresa Molés-Cases & Ulrike Oster</i>	201
Enhancing translator trainees' awareness of source text interference through use of comparable corpora <i>Josep Marco & Heike van Lawick</i>	225
Using a multimedia corpus of subtitles in translation training: Design and applications of the Veiga corpus <i>Patricia Sotelo</i>	245
Applying data-driven learning to the web <i>Alex Boulton</i>	267
Notes on contributors	297
Publically-available corpus tools and resources discussed in the book	301
Subject Index	305
Author Index	311

Introduction

Data-driven learning in language pedagogy

Agnieszka Leńko-Szymańska & Alex Boulton

University of Warsaw / ATILF-CNRS/University of Lorraine

Corpora, large textual databases in the modern sense, first really became available for systematic analysis of language in the 1960s. Since then they have been growing in size and complexity, as have the tools used to process them. Fifty years later we have at our disposal a plethora of corpora of different kinds: monolingual and multilingual collections (the latter containing either original but comparable or translated texts); small highly specialised corpora of a few tens of thousands of words to large reference or monitor corpora of hundreds of millions or even billions of words; corpora containing just text (plain or annotated) or multimodal corpora aligned with sound or video. Present-day tools also go beyond presenting a word in its immediate contexts (KWIC) and providing a wordlist, and offer complex statistical analyses enabling automatic identification of collocations and grammatical patterns, text keywords or stylistic profiles. It can be safely claimed that since their introduction, corpora have changed for ever many research areas: linguistic analyses in the first place, which now are based on attested examples of language use and account for frequency phenomena, but also applied disciplines such as machine translation, speech recognition or forensic linguistics. They are also used outside linguistics, for example in historical or applied social studies, and are commonplace on many websites for the general public (for search functions in particular). One of the areas in which corpora have had a considerable influence is language pedagogy.

In the discussion of the applications of corpora for language teaching we frequently come across the notion of *affordances* (cf. Hafner & Candlin 2007; Boulton 2012). This notion originally comes from the field of perception psychology and was first proposed by the American psychologist James Gibson (1977, 1979). According to this theory, it is a person's perception of the environment that prompts some course of action. Affordances thus refer to the properties of an object in the environment enabling any kind of activity. However, the theory also implies a relationship between the object and the person, and what he or she can do with that object. For example, a book enables reading but can also be used as

support for a leg of a wobbling table and countless other purposes. The realisation of these affordances depends on the person's ability to read or his or her ingenuity to use the book to support the wobbling table, etc. In other words, an affordance can be defined as any use of an object that a person can perceive. This theory has been taken up in various fields including design, human-computer interaction, ergonomics, and so on. Douglas Arthur 'Don' Norman, a well-known researcher in cognitive science, human-centred design and usability engineering, is credited with the observation that "the value of a well-designed object is when it has such a rich set of affordances that the people who use it can do things with it that the designer never imagined" (1994: n.p.).

Corpora have multiple affordances for language teaching which were not immediately apparent to the pioneering corpus compilers, including Randolph Quirk (1960) or Henry Kučera & W. Nelson Francis (1967), who created the first one-million-word corpora of British and American English respectively. However, while more and more researchers carried out projects aiming at improved corpus-based linguistic description (e.g. John Sinclair, the director of the COBUILD project at the University of Birmingham), the implications of corpus analyses for language teaching started slowly to emerge. In 1994 the first conference on Teaching and Language Corpora (TaLC) was held at the University of Lancaster. Its aim was to bring together researchers and language teachers who were involved in pedagogical activities making use of corpora. The proceedings from this conference (Wichmann et al. 1997) were the first published attempt to tackle explicitly and in a wide-ranging way the use of corpora in language education.

In the first chapter of that volume, Geoffrey Leech (1997) proposed a distinction between *direct* and *indirect* uses of corpora in language education. The indirect uses involve situations in which learners come into contact with the results of corpus investigations through the mediation of reference works and teaching materials; though corpora inform such products, the end-user need not even be aware of this. The direct uses of corpora, on the other hand, involve students exploiting corpus data themselves to learn to perform linguistic analyses and, more importantly, to learn the language. Tim Johns (1991: 30), one of the most influential pioneers of the approach, advocated "cut[ting] out the middleman as far as possible and... giv[ing] the learner direct access to the data." He called this kind of activity *classroom concordancing*, the title of the seminal collection of papers he co-edited with Philip King in 1991, and also coined the term *data-driven learning* (DDL) as students were supposed to learn directly from linguistic data, following the motto: "Every student a Sherlock Holmes" (1997: 101). According to Leech, a data-driven approach to learning should be a central application of corpora in education, not least since language itself and language learning *a fortiori* are themselves usage based (e.g. Beckner et al. 2009).

Since the early days of the first TaLC conference, corpora have been increasingly used in language pedagogy. A personal account of the growing influence of linguistic databases in language teaching and systematic evidence to support this observation is offered by Tribble (this volume). Over the last twenty years or so, numerous books, articles and conference presentations have discussed a variety of corpus affordances in education, from more accurate corpus-based descriptions of the target and learner language, to the creation of new corpus-based resources and tools for language teaching and learning (e.g. Ghadessy et al. 2001; Sinclair 2004; Braun et al. 2006; Aijmer 2009; Campoy-Cubillo et al. 2010; Boulton et al. 2012, to name just a small sample of collected volumes on these topics). The academic debate has been followed by real-life pedagogical endeavours. Currently, corpora are present in the production of almost all learner dictionaries (e.g. the *Macmillan English Dictionary for Advanced Learners* – Rundell 2007), reference grammars (e.g. the *Longman Grammar of Spoken and Written English* – Biber et al. 1999; the *Cambridge Grammar of English* – Carter & McCarthy 2006) and usage manuals (e.g. *Practical English Usage* – Swan 2005), as well as in designing courses (e.g. Lee & Swales 2006), producing course manuals (e.g. the *Touchstone* series – McCarthy et al. 2005) and developing supplementary teaching materials (e.g. the *English Vocabulary in Use* series – McCarthy & O'Dell 2005; *Real Grammar: A Corpus-based Approach to English* – Conrad & Biber 2009).

However, the direct uses of corpora in language teaching are treated rather marginally in the literature in the field. This sadly reflects real classroom practice, in particular outside higher education, where explicit data-driven techniques are rarely incorporated into regular teaching procedures, as observed by a number of researchers (Mukherjee 2004; Römer 2009, 2010; Tribble, this volume). Despite the enthusiasm of a handful of specialists who point to the numerous and varied advantages of the approach, language learners rarely have hands-on experience with corpora in mainstream education, at least in prototypical DDL form. This is regrettable because learners can benefit tremendously from the direct use of corpora. They gain access to authentic language, which they can query in a variety of ways for the information which is interesting and relevant to them at a particular moment and which allows them to refine their understanding of how language really behaves. Frequently, the information available in a corpus is richer than any account offered by reference and teaching materials. Moreover, the very act of formulating an appropriate query and making sense of its results contributes to the development not only of their linguistic awareness but also of their general learning skills and autonomy, which are at the forefront of the recent educational movements (Lamb & Reinders 2007).

A very good reason for teachers' reluctance to introduce corpus-based activities into their work (assuming they are aware of them) is the fact that DDL is

perceived to be a demanding and challenging instructional technique. It requires from learners good technical skills in manipulating corpus-analysis software and formulating appropriate queries. These skills need time and effort to develop, and many teachers feel that they do not have space for this in their already-packed curricula. In addition, corpus data are often criticised as being unsuitable for lower-proficiency students, due to their linguistic difficulty and the limited relevance of topics covered (cf. Breyer 2009; Chujo et al., this volume). And since the majority of language learners worldwide are below the advanced stage of proficiency, their teachers may not find it appropriate to expose them to corpus data containing large quantities of authentic language. Still another hurdle to DDL is its perceived incompatibility with the prevalent communicative language teaching methodology, which advocates focus on fluency rather than accuracy, with priority given to language skills and functions rather than form (Widdowson 1990). DDL, on the other hand, promotes focus on language patterns which are frequently, if not exclusively, equated with grammatical structures. Perhaps the most important obstacle is that DDL implies an inductive approach to learning (as discussed in detail by Flowerdew, this volume). In order for DDL to be a valuable learning experience, learners should be able to take an active attitude to language analysis and interpret the results of corpus queries on their own. Such autonomy, although widely advocated by recent educational trends, is not always in tune with institutions' educational traditions or individual students' learning styles (see Boulton 2009; Flowerdew, this volume).

Nevertheless, these limitations can easily be reduced, as suggested by several researchers, including the authors of the chapters in this volume. The technical challenges are addressed by a suggestion to replace students' access to corpus analysis software with printouts which still enable learners to deal with relevant language data for independent analysis but "takes the computer out of equation" (Boulton 2010); this goes right back to Johns (1991), and is reprised in Curado Fuentes (this volume). Other solutions advanced here include replacing standard corpus analysis software with purpose-built learning environments (cf. Chujo et al., this volume; Babych, this volume) or engaging learners in DDL through the use of web-search engines – tools which they most certainly are familiar with (Boulton, this volume).

The inappropriate nature of corpus data for lower-level students is also challenged by the papers in this book. The creation of level-suitable or simplified corpora can be one way of overcoming the obstacle (Chujo et al., this volume); another is building corpora containing text types and topics which are familiar to students from their first language, for example news items. In such scenarios even pre-intermediate learners can benefit from working with corpora (Curado Fuentes, this volume). In some special pedagogical contexts corpora can be used

with learners who have very limited command of a new language, provided that they can draw on their knowledge of a cognate language (Babych, this volume).

The alleged divergence of communicative language teaching and data-driven learning has largely been invalidated over the years, by researchers both within DDL and further afield. Recent developments within the communicative approach have demonstrated that *focus on form* is indispensable in language learning, especially if learners are to reach higher levels of proficiency (Doughty & Williams 1998; Norris & Ortega 2000). However, form should always be studied in context so that students can associate it with meaning and use. This is exactly what DDL offers (Gaskell & Cobb 2004). For example, by studying words in concordance lines, learners can observe their characteristic phraseologies (Thomas, this volume) or pragmatic functions of frequent lexical bundles (Aston, this volume). Yet DDL activities can also help to develop language skills, one of the priorities of communicative language teaching, in particular writing (Charles, this volume) and reading (Babych, this volume; Curado Fuentes, this volume).

The cognitive demands of DDL can be solved by using corpus data more proactively, as suggested again by Johns (1991). This means that corpus data can be used in a more deductive approach to teaching whereby corpus data is not used to elicit a rule or pattern but to exemplify it. Yet another way, proposed by Molés-Cases and Oster (this volume), is for students to work on highly structured and collaborative corpus-based tasks which can provide greater scaffolding and enable them to gain more confidence from working in a group.

However, it has to be acknowledged that problems with incorporating corpora into language instruction can originate in teachers themselves who frequently complain about a lack of guidance on how to take advantage of advances in corpus-based technologies (Heather & Helt 2012; Leńko-Szymańska 2014, in press). As discussed above, an affordance of an object (a corpus in this case) is not inherent in the object itself, but depends on the perception and ability of the user. There has thus been a genuine need for publications reporting on various corpus-related projects around the world, promoting new corpus resources and tools, and demonstrating how to apply them not only in language instruction, but also in the professional training of applied language specialists such as translators (Marco & van Lawick, this volume; Sotelo, this volume).

One forum for promoting various affordances of DDL have been offered by TaLC conferences, which, after the success of the first meeting in Lancaster in 1994, have been taking place every other year, originally in the UK (Lancaster again in 1996, Oxford in 1998) and then in Western Europe (Graz, Austria, in 2000; Bertinoro, Italy, in 2002; Granada, Spain, in 2004; Paris, France, in 2006; and Lisbon, Portugal, in 2008). In 2010 TaLC moved further afield to Central Europe with the conference organised in Brno, Czech Republic; and the University of

Warsaw, Poland, had the privilege to host the 10th edition of the conference in 2012. Each of these conferences resulted in the publication of a volume of selected papers (Wichmann et al. 1997; Botley et al. 1996; Burnard & McEnery 2000; Kettemann & Marko 2002; Aston et al. 2004; Hidalgo et al. 2007; Kübler 2011; Frankenberg-Garcia et al. 2011; Thomas & Boulton 2012), each with important sections devoted to reporting a variety of DDL projects around the world.

The papers in this volume arise from the TaLC10 Conference in Warsaw; all those selected for publication here have a specific focus on one particular affordance of corpora in language pedagogy – data-driven learning. This focus on corpus use by learners intends to fill a gap in the literature, as identified for example in a recent review of a book in the same *Studies in Corpus Linguistics* series in which Cheong (2013: 2) regrets there are not “more articles focused on pedagogical implications of how to use corpus-informed research for classroom uses.” By concentrating on learners’ own use of corpus tools and techniques, and by providing accounts of the authors’ first-hand experiences from such projects, this volume addresses the need for relevance to various language teaching practitioners.

The book comprises a collection of 12 papers by corpus linguists and applied linguists, all with considerable experience in language and translation teaching and research in higher education. Its aim is to present recent advances in DDL, addressing issues involving different types of corpora, for different learner profiles (lower-level learners, specialist and non-specialist students, etc.), in different ways for different purposes (exposing students to concordance lines in their native language, finding solutions to the cognitive and technical demands), and using a variety of different research methodologies and perspectives.

The first two chapters set the stage for more specific accounts of DDL projects covered in this book by discussing the theoretical underpinnings of corpus-driven language pedagogy and its effectiveness, as well as by presenting the development of DDL from its conception to the present date, discussing the range of current uses and possible future lines of development.

Lynne Flowerdew (Chapter 2) provides a theoretical foundation for the applications of corpus data described in the volume by discussing several language learning theories underpinning DDL: the noticing hypothesis, constructivist learning and Vygotskian sociocultural theories – theories which, remarkably, are rarely discussed in depth or made explicit in accounts of DDL projects. Thus, this paper illustrates how the tenets of the three language learning theories support and are realized in DDL, drawing on examples of key studies in this field. Flowerdew reports on the small number of existing experimental investigations which address questions of the effectiveness and affordances of DDL by examining learners’ *performance* (rather than attitudes or evaluations) in reference to these three theories, as well as in relation to the concept of learning styles. The conclusion she draws

from the review of these studies is that while DDL shows promise, a larger number of large-scale and longitudinal studies are needed to confirm its beneficial effects.

In Chapter 3, **Chris Tribble** provides an overview of the history of the application of computers and corpora in language education from a personal perspective. He remarks on how the keystones in his career as an applied and corpus linguist were marked by rapid changes in technology and the growing availability and functionality of hardware, software and textual databases. The second part of the paper reports on the results of a questionnaire distributed on various Internet-based discussion lists in 2012 addressing the practices of language teachers in relation to DDL. The questionnaires provide information on the teachers' and their students' profiles, their reasons for using or not using DDL in their teaching, the resources and books used and found useful in the area, and respondents' perceptions of the benefits and disadvantages of DDL tools. Comparing the results with a similar questionnaire circulated four years earlier, Tribble concludes that the results are encouraging given the "emerging maturity of corpus-informed language teaching", and he offers predictions for future developments in this area.

Each of the 9 chapters in the main body of the book describes a project designed to place learners or teachers in direct contact with corpora, and encourage them to analyse corpus data for a variety of purposes. The papers are grouped into three parts: Corpora for Language Learning (Chapters 4, 5 and 6); Corpora for Skills Development (Chapters 7, 8 and 9); and Corpora for Translation Training (Chapters 10, 11 and 12).

Part I examines how corpora can help learners to improve their language knowledge, describing projects where learners exploit corpora as an aid to gaining better control over linguistic systems (phraseology and grammar of the target language).

The focus of the first chapter in this part (Chapter 4) is on using DDL techniques for teaching foreign language phraseology to future interpreters. **Guy Aston** subscribes to the view that phraseology is key to fluency in both speech production and reception, particularly in cognitively and affectively demanding contexts like interpreting. The originality of this project lies in the use of a speech corpus for this purpose: a one million-word collection of talks from the TED – Ideas Worth Spreading website which aligns transcripts with audio/video files. The activities proposed in this chapter aim not only to enrich learners' limited repertoires of phraseological items, but also to improve their command of the associated lexicogrammatical, functional, and prosodic aspects – the last of these being impossible when exploiting transcriptions alone without access to audio material. The examples discussed in the chapter provide some ideas as to how phraseological items and their relevant linguistic properties can be identified, interpreted and made salient through DDL.

A similar claim about the importance of phraseology is made by **James Thomas** (Chapter 5), who advocates the use of full texts as a starting point for tackling phraseology. This allows learners to observe collocations associated with salient words in the text, and the typical meanings and uses of these collocations can be further explored in a corpus. However, Thomas argues that DDL cannot stop there as such activities do not guarantee active knowledge or appropriate use of observed collocations. He proposes *Collocation Plus* as a procedure whereby learners explore the lexical and grammatical environments of collocations in the contexts in which they meet them in corpora, and then record their findings in 'word templates' which will later be available for production. Thomas argues that this procedure will help learners produce more accurate, fluent, idiomatic and sophisticated language, whilst developing their autonomy in using available corpus resources and tools.

The last chapter of this part (Chapter 6) addresses the need for corpora and corpus tools which are suitable for low-proficiency learners of English as a foreign language (EFL). **Kiyomi Chujo**, **Kathryn Oghigian** and **Shiro Akasegawa** describe the creation of two new resources designed especially for such students in Japan in order to help them improve their knowledge of grammatical structures which they find particularly problematic. The first resource is a database of example sentences called the Sentence Corpus of Remedial English (SCoRE), which contains sentences featuring target grammatical structures. Although the sentences are specially written by the researchers, their creation was preceded by a careful analysis of lexical and structural characteristics of a 30-million-word corpus comprising various authentic educational materials relevant for lower-level learners. The second resource is a new DDL tool, the Grammatical Pattern Profiling System, which is a browsing program with a simple interface in which the results appear as complete sentences rather than in KWIC format. Teachers and students can search the database by grammatical category or target word to retrieve examples of a target structure from SCoRE. The authors hope that the browsing system will promote DDL among lower-level learners.

Part II concentrates on the development of language skills. The projects described here contain DDL activities whose aim is to give learners the opportunity of becoming more fluent language users by developing their skills in reading and writing.

Maggie Charles (Chapter 7) advocates the use of self-compiled specialized corpora by advanced learners of English for academic purposes (EAP). Her paper presents an approach to implementing DDL in a multidisciplinary EAP course with the goal of helping students improve their writing skills in their individual academic disciplines. In class students work with their own corpora of academic articles from their own disciplines, but they all complete the same teacher-guided

tasks, designed to investigate key features of written academic discourse. The advantage of this approach is that students' discoveries of the features characteristic of their own disciplines are enhanced by the comparison with other students' discipline-specific results. The paper presents the use of several corpus tools (concordancer, collocation finder, wordlist and distribution plot) for exploiting EAP corpora and suggests that all have valuable pedagogical applications that go beyond the traditional KWIC format of DDL. The author concludes by suggesting that the opportunity to explore different sets of data and to compare the results brings an additional benefit to DDL.

In the case of cognate languages, DDL can help learners use an L2 as a springboard to an L3, as demonstrated by **Svitlana Babych** (Chapter 8) for reading skills. This chapter proposes a methodology which can help students develop top-down processing strategies, in particular the awareness of text organization, by focusing on textual connectors in the L1, L2 and L3. This methodology has been embedded in an on-line environment to support text comprehension in L3 Ukrainian and L2 Russian, which incorporates two elements. One is a comparable corpus of online newspaper articles in English, Russian and Ukrainian, providing a source of texts of similar type or genre for teaching reading. The other is a multilingual electronic thesaurus of textual connectors for the same three languages, systematically derived from large corpora and classified according to Babych's own semantic classification scheme. Babych discusses the ways in which the environment can be used for developing reading skills, and suggests that although this methodology was developed primarily for specific languages, it can be applied to the development of both reading and writing skills in any L2 as well as an L3 which is closely related to the L2.

Alejandro Curado Fuentes (Chapter 9) also describes an application of DDL to developing reading comprehension of news articles. However, this chapter focuses on exploiting news keywords in various DDL activities which aim to raise students' awareness of linguistic properties of the target words, such as parts of speech, collocations and colligations. This is based on the assumption that exploiting news keywords through various DDL tasks can aid learners in subsequent decoding of information in a text. Curado Fuentes conducts an experimental study with business English students at B1 level in which DDL is compared with more conventional paper-based activities. Pre-tests and post-tests were used to compare the progress of experimental and control groups, while direct observation, logs and questionnaires provide further insight. The results indicate that DDL activities focusing on news keywords work favourably for the students' comprehension of news articles.

Part III presents various ways in which corpora can be used in translation training. It describes data-driven tasks promoting general language awareness as