

Zeitschrift: Bulletin suisse de linguistique appliquée / VALS-ASLA
Herausgeber: Vereinigung für Angewandte Linguistik in der Schweiz = Association suisse de linguistique appliquée
Band: - (2013)
Heft: 97: Data-driven approaches to language learning : theories, methodes, applications and perspectives = Apprentissages sur corpus : théories, méthodes, applications et perspectives

Artikel: Grammatical description and classroom application : theory and practice in data-driven learning
Autor: Kerr, Betsy
DOI: <https://doi.org/10.5169/seals-978585>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 09.02.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Grammatical description and classroom application. Theory and practice in data-driven learning

Betsy KERR

University of Minnesota

9 Pleasant St. S.E., Minneapolis MN 55422, USA

bjkerr@umn.edu

Partant de certains concepts intéressants du domaine de la linguistique sur corpus, cet article défend l'idée de faire une place plus importante à l'acquisition du vocabulaire et de mieux intégrer les aspects lexicaux et grammaticaux dans l'enseignement de la langue. Nous nous intéressons tout particulièrement à l'idée de la centralité des séquences lexicalisées dans les productions orales, telle qu'elles sont identifiées dans la grammaire lexicale de Sinclair puis appliquées à la didactique des langues avec l'approche de Lewis. De plus, cet article montre, à partir d'activités types, comment l'apprentissage à partir de corpus est propice aux activités métacognitives centrées sur l'apprenant, pointant des aspects importants concernant le lexique et la grammaire du français. Finalement, nous passons en revue plusieurs outils qui se prêtent à la préparation de ce type d'activité par l'enseignant.

Mots-clés: Concordance, corpus, approche lexicale, didactique des langues, linguistique sur corpus, FLE, concordanciers en ligne

1. Introduction

Many proponents of data-driven learning (DDL) or the use of concordance-generated data in language teaching and learning have lamented the fact that second language educators, especially outside of the field of English L2 instruction, have not embraced the use of DDL in significant numbers (see for example Cheng 2010: 327-328). My purpose here is to encourage L2 French teachers in particular (and those responsible for L2 French teacher education) to consider trying out concordance-based activities in their classrooms. I will begin by seconding others' suggestions for a progressive approach to developing learners' abilities to utilize DDL activities, as well as enumerating some of the holistic benefits that can accrue from the regular practice of DDL. The next section will sketch out Sinclair's idiom principle, a view of language originating in corpus linguistics which sees lexis and grammar as intimately related. A pedagogical application of this view, Lewis's Lexical Approach, will be briefly described as a model for a pedagogical approach in which DDL plays a central role in facilitating learner discovery of patterns of language use that integrate lexical and grammatical properties.

The third and most substantial section of the article will be devoted to practical questions relating to teachers' use of concordance-based or DDL activities. Following a presentation of types of language features that can be addressed by means of DDL, concrete suggestions are given for transforming

concordance data into classroom activities, and a number of examples of such activities are provided. Finally, four online French corpora with built-in concordancers that lend themselves to teacher and student use are described and compared.

2. Benefits of teacher-mediated DDL activities

Boulton (2011) proposes a prototype definition of DDL situating pedagogical uses of corpora consultation on a continuum of weak to strong versions of DDL, according to the number of prototype criteria exhibited by the given application¹. Pursuant to this framework, activities other than hands-on use of raw concordance data, such as teacher-devised classroom activities based on edited concordances, constitute 'weaker' versions. Nevertheless, in Boulton's opinion and as demonstrated by explicit recommendations of the originator of the term (Johns [1990] 1991, 1993), such activities have a rightful place and play an important role in the panoply of corpus-based learner activities (see Boulton & Tyne this volume).

It is precisely activities at this end of the continuum, i.e. teacher-mediated encounters with corpus data, that will form the object of this discussion. The motivation for this choice is the belief that teacher-mediated activities offer language teachers the easiest entry into DDL and, in turn, offer the best chance for the expansion of the application of DDL in L2 instruction. A typical activity of this type engages learners in a teacher-structured process of examining and analyzing concordance data which has been assembled and edited by the teacher with a particular group of learners and particular learning goals in mind. Though such teacher interventions may diminish certain aspects of the learner's experience that have been touted as beneficial (e.g. the encounter with totally authentic language and the autonomous nature of the learner's role in formulating hypotheses), these modifications have the advantage of making corpus-based work accessible to a larger student audience than would otherwise be possible. Obviously, teacher-produced concordances do not require the same technological resources as hands-on work, and, more importantly, if well executed, the teacher's adaptations of the data will facilitate its interpretation by learners with less advanced linguistic skills than are required to decipher unedited concordancer output, with its partial sentences, limited context, and unpredictable and ever-changing topics. Though these features will remain in any concordance data, judicious selection of concordance lines, as well as the provision of guidance in the process, can make these challenges considerably more manageable for learners of less advanced proficiency levels.

¹ Gabrielatos (2005: 11) provides a similar characterization and examples of concordance activities of a "soft" version (teacher-controlled) and a "hard" version (learner-controlled).

Teacher mediation of corpus investigation thus has the potential to blunt one of the major criticisms of DDL, namely, that it can be effective only with learners of relatively well-developed linguistic skills and / or high levels of motivation (Cheng 2010: 328). Another common complaint (Chambers 2010: 348), that corpus consultation is time-consuming and laborious, is countered by the streamlining afforded by the teacher's crafting of both the data for analysis and the task itself. Several proponents of DDL (Boulton 2011; Chambers 2010) have put forth a progressive approach, starting from maximally teacher-structured and teacher-aided activities and moving gradually toward independent, unaided use of concordances. Ideally, this would be a long-term process extending over as much as two years or more of study and coinciding with the learners' advancement in degree of competence in the language. Chambers (2010: 349) proposes a three-phase model as follows: (1) working with teacher-prepared concordance printouts to become familiar with the inductive process, (2) consultation of user-friendly online resources, and (3) independent corpus consultation selecting from a range of corpora. The hope is that this kind of long-term, recurrent practice will result in increased voluntary recourse to concordance data as a reference and writing aid on the part of advanced learners. Moreover, this approach provides a "gentle 'way in'" (Johns 1993, as cited in Boulton 2011: 573) not only for learners, but also for teachers, who have so far not embraced DDL activities in significant numbers.

Proponents of DDL argue that investigation of corpus data is a more effective learning tool than the conventional deductive approach of presenting rules with a few examples, thanks primarily to two features of DDL, namely, the engagement of the learner in discovering patterns for themselves (see 3.3 below) and the provision of a larger number of examples. In addition, the authenticity of the examples is believed to enhance learning. However, in addition to affording a more effective means of acquiring particular patterns of language use, DDL activities, when practiced with some regularity, offer more holistic benefits that may be just as important, or even more important, than the knowledge gained of specific points of vocabulary or grammar. O'Sullivan (2007: 278) cites numerous researchers who claim that engaging in DDL activities has the potential to bring general cognitive benefits which are vital to language learning such as the development of analytical skills, the ability to notice linguistic features, and problem-solving abilities (see also Landure this volume). Stevens (1991: 48) adds to these "the ability to extrapolate holistically from fragmentary evidence". My own observation is that most learners do not come to language study with a syntactic mind-set; that is, L2 learners, especially in the case where L1 and L2 share a great deal of common syntactic structure and lexis, naturally tend to focus, in their efforts to extract meaning from L2 texts, on meaning rather than on form. Consequently, when learners then come to attempt to express meaning in writing or

speaking, their awareness of morphological and syntactic features is insufficient to allow them to formulate language which is native-like in its formal features. By engaging learners in activities that require attention to formal features such as grammatical functions of sentence constituents, part of speech of lexical units, and function words (prepositions) introducing complements, teachers can lead learners to acquire the habit of paying attention to these features, while stressing the importance of awareness of such features in producing effective and accurate speech and writing. In addition, we will see below examples of the generalizability of certain patterns from a particular syntactic structure or lexical item to a larger category of linguistic items.

3. Theory and application: the relationship between lexis and grammar

3.1 Sinclair's idiom principle

Second language instruction, regardless of the context or the approach adopted (and even in the context of widely adopted communicative approaches), has over recent decades privileged explicit instruction in grammar, i.e. morphological and syntactic patterns of the L2. On the other hand, acquisition of lexis has generally been left to learning of the incidental type, believed to take place more or less naturally, through reading or other encounters with the L2, thanks to redundancy and contextual clues present in natural texts and communicative interactions. While many studies (e.g. Hulstijn *et al.* 1996, Pigada & Schmitt 2006) have demonstrated positive albeit somewhat limited effects of extensive reading on vocabulary acquisition, incidental learning of lexis is usually restricted by the limited time learners actually spend in reading or other naturalistic interactions in the target language (Sockett 2011).

Insights gained over the past few decades from studies employing corpus-based methods, as well as from other fields of investigation, have created a new interest in the structure of the lexis and its role in L2 acquisition. In particular, corpus data have revealed the extent to which certain combinations of words occur with greater frequency, even where other combinations would be just as valid from a syntactic and semantic point of view. Much work has been devoted to these frequent combinations or collocations, which have now become the basis for the leading learner-oriented dictionaries as well as some course materials, first for English and then for the other most-studied European languages².

² Pioneers in this field were the Cobuild group led by John Sinclair which published the first *Collins COBUILD English Language Dictionary* in 1987 (5th edition 2006) and the *Collins COBUILD English course* (Willis & Willis 1989).

John Sinclair, a pioneer of corpus linguistics with an interest in its application to second language acquisition, was so impressed by the prevalence of collocations and other multi-word units in actual native-speaker language that he called into question the very distinction between lexis and grammar. Accordingly, he proposed the notion of a "lexical grammar" where the grammar and lexis are built together, with each on an equal basis (Sinclair 2000: 191). To better understand the kind of patterns motivating this proposal, let us consider an example of lexical dependencies revealed by corpus data. As reported by Hunston and Francis (2000: 24-25), in a 1994 paper³, Sinclair develops the example of the phrase *naked eye*. Inspection of even a short concordance⁴ leads to the following description of the phrase's typical contexts:

Typically, *naked eye* occurs at the end of a clause and is preceded by *the*. Furthermore, *the naked eye* is preceded by *to*, or, less frequently, *with*. The prepositional phrase *to / with the naked eye* follows a range of words related to sight, most frequently *visible* or a form of the verb *see*. Prior to that is an indication of something that might be too small to be seen. About half the instances of this typical usage are negative (as against a general figure of 10% of all clauses being negative, see Halliday 1993b); in other words, something is described as being *invisible to the naked eye*. Of the positive instances, over half include modification, such as *easily*, *actually*, *might be* or *the first*.

Sinclair concludes from these observations that *naked eye* is part of a larger 'meaning unit', which has a consistent meaning, even as it displays a range of possible lexical realizations.

It is this kind of observation that leads Sinclair to articulate the idiom principle: semi-preconstructed phrases such as collocations actually represent a single choice made by the language user, despite their appearing to be analyzable into segments (Sinclair 1991, cited in Hunston & Francis 2000: 21). This process of speech production through the choice of language chunks is thus opposed to the conventional 'slot-and-filler' model, where "the syntactic structures form a series of slots, and these are filled with choices from the dictionary", the lexical items being chosen individually, with relatively little reference to the surrounding text (Sinclair 2000: 195).

Sinclair's work in corpus linguistics is paralleled by research in related fields such as second language acquisition and teaching, and psycholinguistics. At the same time, the object of study has been expanded to include various types of multi-word units, generally distinguished by their relative degree of fixedness. Schmitt (2000: 99-102) provides a summary of some of the proposed taxonomies, singling out as more common the categories of "compound words, phrasal verbs, fixed phrases, idioms, proverbs, and lexical

³ "A search for meaningful units of language", paper read at the International Symposium on Phraseology, University of Leeds, April 1994, as reported in Hunston and Francis (2000: 24-25).

⁴ I recommend the Time Magazine Corpus created by Mark Davies at Brigham Young University: <http://corpus.byu.edu/time/> (17.05.2013).

phrases". Terms employed by other researchers for lexical phrases include routine formulae, gambits and lexical chunks (Schmitt 2000: 101), and the study of these units is generally referred to as 'phraseology'. Among these researchers, the lexicographical perspective is represented by Mel'čuk's work on a taxonomy of "non-free phrases" (Hunston & Francis 2000: 8-10). In psycholinguistics, lexical phrases have figured prominently in debates about L1 and L2 acquisition, as well as about how language is encoded in the brain, whether as single lexical items to be combined with the help of grammatical rules, or as multi-word units. For example, Peters (1983) argues that L1 acquisition proceeds first through the acquisition of lexical phrases that are later analyzed into a system of words and structure (Hunston & Francis 2000: 11-12). Conklin and Schmitt (2008) provide support for this hypothesis by demonstrating a processing advantage in the reading of formulaic vs. non-formulaic sequences on the part of both native and non-native readers. In the field of language teaching, Pawley and Syder (1983) and Nattinger and DeCarrico (1989, 1992) highlight the role of the acquisition of lexical phrases in the achievement of fluency: given that lexical phrases are stored as whole units, their retrieval saves much-needed processing time, as compared to that required for on-line processing of several items which must be individually selected and sequenced according to grammatical constraints.

3.2 Pedagogical application: Lewis's Lexical Approach

Michael Lewis's Lexical Approach (1993) and Lewis (1997) describe an approach to second language teaching which Lewis believes to be consistent with a lexically based view of language such as that outlined above. In what follows, I will highlight some of Lewis's basic principles and important methodological recommendations, especially as they pertain to grammar instruction and to the use of corpora and concordancing. The following are among the more distinctive "key principles" laid out by Lewis (1993: vi-vii):

- Language consists of grammaticalized lexis, not lexicalized grammar.
- The grammar / vocabulary dichotomy is invalid; much language consists of multi-word 'chunks'.
- A central element of language teaching is raising students' awareness of multi-word units, and developing their ability to 'chunk' language successfully.
- It is the co-textual rather than situational elements of context which are of primary importance for language teaching.
- Grammar as a receptive skill, involving the perception of similarity and difference, is prioritized.
- The Present-Practice-Produce paradigm is rejected, in favor of a paradigm based on the Observe-Hypothesize-Experiment cycle.

- ... the materials and methods appropriate to beginner or elementary students are radically different from those employed for upper-intermediate or advanced students.

The Lexical Approach seeks to direct learners' attention to what are "highly probable" utterances, not simply to all possible utterances, as Lewis suggests is the case with typical grammar books, following the lead of generative grammar (1993: 12).

Moudraia (2001: 2) gives the following summary of the kinds of activities proposed by Lewis to develop learners' knowledge of multi-word units:

- Intensive and extensive listening and reading in the target language.
- First and second language comparisons and translation – carried out chunk-for-chunk, rather than word-for-word – aimed at raising language awareness.
- Repetition and recycling of activities, such as summarizing a text orally one day and again a few days later to keep words and expressions that have been learned active.
- Guessing the meaning of vocabulary items from context.
- Noticing and recording language patterns and collocations.
- Working with dictionaries and other reference tools.
- Working with language corpora created by the teacher for use in the classroom or accessible on the Internet ...

Despite Lewis's de-emphasis of grammar in favor of lexis, he does acknowledge a place in the language syllabus for grammar, rightly understood. Seeing grammar as "the search for powerful patterns" (1993: 137), Lewis suggests certain aspects of grammar which deserve increased attention. Two of these are word grammar, the set of patterns in which the word occurs, and de-lexicalized words. For words with relatively high meaning-content (such as *deposit*, *house*, *baby*), Lewis suggests the following pedagogical strategies: identifying high-frequency uses and distinguishing among them; identifying cognates, that is, derived or morphologically related forms, and their collocations; identifying and recording high frequency collocations, any natural noun phrase – verb phrase combinations, and any other lexical phrases in which the word occurs (1993: 142). De-lexicalized words are words of low meaning-content, which allows them to serve as "powerful pattern generators" (1993: 143). Examples for French are the verbs *faire*, *avoir*, *prendre* and function words such as the prepositions *de* (in at least certain uses) and *en*. Clearly, these are the kinds of patterns that can be explored through corpora and concordances.

3.3 Constructivist learning theory

The notion of learner as scientist, implicit in Lewis's Observe-Hypothesize-Experiment cycle, draws support from the branch of learning theory known as constructivism. This model has recently attracted increased interest from educational technologists, as developments in informational technology have greatly expanded the quantity and ease of use of data sources available for data-driven learning activities. Cobb (1999: 2)⁵ contrasts the two types of knowledge believed to be created by the traditional transmission of knowledge vs. the construction of knowledge by the learner:

One kind is inert, easily forgotten, and untransferable beyond its initial context of learning because it was "pre-emptively encoded" (Spiro, Coulson, Feltovich & Anderson 1988: 377) by an expert for transmission to a learner. The other kind is memorable and transferable to novel contexts because learners have encoded it for themselves out of raw data, or at least raised it from a lower to a higher level of organization, by forming and testing hypotheses as professional scientists do (Resnick 1987).

In the same article, Cobb (1999: 1) reports on a study in which he devised and carried out an empirical test of the prediction that knowledge of the second type would be more "flexible, transferable, and useful" than knowledge of the first type. Subjects in the study were students learning English for academic purposes who were engaged in two different sets of activities for vocabulary expansion: learning dictionary definitions of words for the control group vs. constructing their own definitions of words based on observation of the word in several sentence contexts drawn from concordance data. These latter activities were facilitated by the use of a specially designed 'dictionary toolkit' modeled on a lexicographer's workstation, which linked together a teacher-generated wordlist and a learner-created database with simplified versions of a corpus and concordance engine. The study results showed that the group using concordance data to construct definitions attained better results than the control group in tasks requiring the transfer of their word knowledge to novel contexts.

4. Pedagogical uses of concordancing

4.1 Teacher manipulation of the concordance and the task

We turn now to instructional activities where the teacher mediates between the raw data and the learner, reducing the complexity of the task but retaining the essential elements of DDL. Chambers (2010: 348-349) cites two main objections voiced by learners who have tried out DDL activities: first, the "time-consuming, laborious and tedious" nature of the activities, and second, the limited nature of the context supplied by concordance lines. The latter difficulty

⁵ Citations are from the author's pre-publication version:
<http://www.lex tutor.ca/cv/Constructivism.html> (01.09.2007).

can be abated to some extent by choosing from available corpora those with text-types most likely to be familiar to the given group of students; however, in most cases, learners will have to grapple with limited context. To address the first objection, teachers will reduce the raw concordance to a manageable size by selecting a limited number of lines that lend themselves to the objective of the activity. In addition, learners' interpretation of concordance lines can be significantly aided by the teacher's judicious selection of concordance lines according to their anticipated readability by the given group of learners (Gilquin & Granger 2010). Finally, the activity can be made less tedious and more efficient by the teacher's construction of the task and provision of clear guidelines for the learners to follow.

4.2 Taxonomy of linguistic features for corpus-based study

Before looking at some examples, it will be useful to review the principal kinds of patterns that can be studied with the aid of a corpus and concordancer. The following inventory of features is loosely based on Aston's discussion of pedagogical concordancing (2001: 16-19). Of course, the particular features that teachers single out for special attention should ideally be selected from texts with which learners are working, or based on recurrent errors or problem areas for the given group of learners. Also, the scope of any activity will be determined by the level and needs of the learners.

1. Lexical items (words)
 - polysemous words: different meanings of the word and the contextual patternings associated with each of them. Examples: different uses of *tirer*, *projeter*, *abattu*, *passer*.
 - lexico-grammatical features, or word grammar: different syntactic contexts and the meanings associated with them. Examples: *réussir* + noun vs. *réussir à* + noun / infinitive; reflexive vs. Non-reflexive uses of *sentir*; *penser à* + noun / infinitive vs. *penser* + infinitive, etc.
2. Multi-word units (set expressions, collocations, etc.). Examples:
 - phrases with *même*, e.g. *même si*, *quand même*, *tout de même*, *quand bien même* *que*, *de même* (*que*).
 - phrasal verbs such as *mettre en* + NP (*œuvre*, *doute*, *garde*, *place*, etc.).
 - idiomatic expressions, especially those which may be absent or difficult to locate in dictionaries: *montrer patte blanche*, *chercher midi à quatorze heures*, *tenir la dragée haute à quelqu'un*.
3. Derivational morphology: prefixes and suffixes. Examples:
 - the prefix *anti-*: to what type(s) of word does it attach, what types of words does it form?

- the suffix *-ment*: to what two types of words does it attach, forming what two types of words?
4. Grammatical patterns. Unless one is working with a tagged corpus (one in which the syntactic category of each word has been indicated in the text so that it can be read by the concordancer), this type of search will be limited to grammatical structures which are consistently represented by, or which regularly co-occur with, a small set of lexical items, such as the following. Examples:
- relative pronouns *qui*, *que*; *dont*; *lequel* and contractions (in untagged corpora, concordances will need to be edited to include only the relevant occurrences of these forms).
 - the adverbial expression *ne ... que* (most concordancers allow searches for discontinuous features, often through the use of regular expressions).
 - subjunctive verb forms: grammatical forms having many different realizations can sometimes be located by searching for associated items. For example, a good number of subjunctive forms may be found by searching for *il faut que*, *à moins que*, or *bien que*. The converse approach, if investigating the various contexts that trigger subjunctive use, would be to search for very frequent subjunctive forms such as the third-person forms of *être* (cf. Chambers this volume).
5. Comparison of items which are distinct but which present some similarities. The differences may be semantic or primarily syntactic. Examples:
- homonyms. Examples: grammatical forms *leur* (pronoun vs. determiner), *lui* (dative pronoun vs. disjunctive pronoun). Similarly, referential vs. non-referential pronouns *il* and *ce*.
 - synonyms (or so-called 'translation synonyms'). Examples: *amener*, *apporter*, *ramener*, *rapporter*, *faire venir*, all related to Eng. 'to bring'; *briser*, *casser*, *rompre*, related to Eng. 'to break'; prepositions *en*, *dans* ('in').
 - paronyms (items of similar form but different meaning, whether related or not). Examples: *accroissement* ('increase') and *croissance* ('growth'); *jouir* and *jouer*; *avoir la chance de* + infinitive vs. *avoir des chances de* + infinitive.
 - prepositions vs. corresponding conjunctions. Examples: *après* and *après que*, *avant* and *avant que*, *depuis* and *depuis que*, etc.
 - *passé composé* and *imparfait* forms of verbs such as *vouloir*, *connaître*, *avoir*, where the aspectual distinction between the two tenses occasions differing translations in English.

6. Comparison of the same feature in different text types.
Examples: comparison of an oral and a written corpus with respect to use of subjunctive, negation, or interrogative forms. In addition, all of the corpora of written French described below allow choices of various genres such as literary, journalistic, scientific, political and administrative; patterns of lexical usage are often quite different depending on the genre examined.

4.3 Designing concordance-based learning tasks

Concordance-based activities lend themselves to a wide variety of task types. Most often, learners are asked to study the concordance data prepared by the teacher and to draw generalizations about either structure or meaning of the targeted item or items. A simple case would employ a short concordance of one lexical item or lexical phrase and ask the students to infer the meaning of the expression. This can be done rather quickly (with the use of a PowerPoint presentation or, if technology is more limited, an overhead transparency) and is an effective means of developing learners' ability to guess meaning from context. This can be a useful technique for clarifying the meaning of unfamiliar expressions in a text that the class is reading together, as the single context of the text in question may not be sufficient to give learners confidence in their guessing. Once the students have formulated a hypothesis on the basis of several contexts, the hypothesis can be verified by reference to a dictionary definition. This kind of activity is facilitated by the use of Cobb's Compleat Lexical Tutor site (see below) which allows the uploading of a text and creates a hypertext providing access to a concordance and dictionary entry for any word of the text.

More complex tasks will require learners to draw generalizations about structure or meaning where the concordance data contains multiple patterns, either structural or semantic or, more often, both. In such cases, the task can be simplified by presorting the data, or alternatively, learners can be asked to sort the concordance lines into categories which are given to them or which they will define. These tasks work well with pairs or small groups of students, who may be asked to write out their findings for comparison with their peers' results in the follow-up discussion, which is of course essential. If the concordance data is somewhat lengthy, different parts of the concordance can be assigned to different groups; in this case, the parts of the concordance may be mixed, displaying various patterns, or they may be designed so as to separate the different patterns, so that each group focuses on one of the patterns only. A pattern-discerning activity can also be developed into a more or less extended worksheet, where the learners respond to a series of questions that lead them through various observations or more complex reasoning.

Another approach to the use of a concordance is to have learners carry out some operation that will engage them in meaningful processing of the concordance lines, thus exposing them to multiple instances of the target item. A typical task of this sort is a matching exercise where the concordance lines have been split at the keyword and one side of the contexts reordered, as in Example 4 of the section below, which requires processing of clauses conjoined by *de sorte que*. This type of exercise is particularly useful for providing practice in making use of contextual clues to construct meaning from a written text. Other exercise types have as their purpose the reinforcement or assessment of the learner's acquisition of a distinction or set of patterns which have previously been the focus of a concordance-based discovery activity. These are typically of the gap-filling type, but employ only authentic concordance data. Römer (2006: 125) presents a gap-fill exercise of this type for collocations with *speak* and *talk*. More creative sentence-completion tasks, where the part of the sentence that is supplied closely resembles part of (one of) the target pattern(s), may also be utilized for this purpose. Another possible approach is to have learners produce their own exercises based on what they find in a set of corpus data.

4.4 Sample concordance-based activities

The following are examples of concordance-based activities that I have used in teaching advanced-level post-secondary American students of French. The use of concordancing is virtually unknown in the U.S. French L2 educational context, which I attribute to two primary causes: the relatively low levels of proficiency attained by most American learners given their late start at L2 education, and the prevailing literary orientation of French L2 curricula at the postsecondary level. These two factors make it difficult to incorporate significant work with corpora in the regular curriculum. However, I have created and used activities such as these to focus on particular points of lexis and grammar that pose difficulties for large numbers of students on a recurring basis. Students have responded positively to these teacher-mediated activities; occasionally, the more linguistically curious among them will ask questions about linguistic aspects not in the focus of the lesson. Activities such as these can be completed in as little as fifteen or twenty minutes of class time. Though this is more time than is required to present an explanation and examples in the traditional fashion, I believe that these activities are more effective because of the fact that each learner is required to grapple with the language samples and formulate their own generalizations.

Due to space limitations, the concordances in the following sample activities are reduced to a bare minimum. Typically, such concordances would include ten to twelve concordance lines or more, but it is of course the teacher's prerogative to choose the optimal length, depending on variables such as the

nature of the task and the number of different usages, meanings or constructions to be included.

Example 1. Single-word polysemous lexical unit: *projeter*.

Task: semantic classification.

Concordance source: Lexiquum (La Presse – see section 5 below).

1. probablement inclus en partie dans les résultats passés de l'équipe. Ceux dont nous nous servons pour **projeter** l'avenir.
2. la robe! Pour le prêt-à-porter, les prix sont plus modestes: de \$1000 à \$2500. Chez Balmain on veut **projeter** l'image de la femme moderne, élégante. On ne présente pas de robe folle, pas de robe marrante.
3. films, 13 fois par mois en moyenne. Chez les jeunes de 18 à 24 ans, le magnétoscope sert surtout à **projeter** des films loués au club vidéo.
4. Une camionnette circulant en direction ouest, sur cette rue, aurait alors happée la voiture pour la **projeter** contre l'autobus de la STCUM qui circulait dans la direction opposée.
5. ment Blake Edwards. Le meilleur moyen de surmonter son trouble, c'est encore de le décrire ou de le **projeter** sur un écran.
6. architecte, Mme Lambert a dit à La Presse, hier soir, tout l'espoir qu'elle a de voir Montréal mieux **projeter** l'image de sa personnalité sur la scène mondiale et mieux se représenter comme collectivité

Discussion: students can be asked to identify the various meanings of the verb exemplified in the data, or they can be given four meanings and asked to classify the examples and identify collocations: 'to throw' (+ locative preposition), the concrete sense 'to project' (*un film, des images*), the figurative sense 'to project' (*une image*), and another figurative extension, 'to expect'. Particularly at the advanced level, discussion can bring attention to the nature of the relationships between the various meanings, i.e. concrete and figurative, relationships which are of course generalizable to many other lexical items.

Example 2. Lexico-grammatical features: *sentir* and *se sentir*.

Task: distinguish and describe two grammatical patterns.

Instructions to student: complete the table below the concordance by writing in the appropriate column the complement that follows each occurrence of *sentir* (non-reflexive use) or *se sentir* (reflexive use). (A *complement* is the word or phrase which accompanies a verb, usually following the verb, and which completes the meaning of the verb.) Then answer the questions following the table.

Concordance source: Lexiquum (Éditions Leméac 1991-1993).

1. Je ne pensais pas qu'elle pouvait **sentir** ma colère.
2. J'ai fini par apprendre qu'elle venait ici pour ne pas **se sentir** trop seule, qu'elle aimait les livres et le silence, ...
3. Des exercices, de la marche, des séances de méditation m'aideraient à **me sentir** plus à l'aise.
4. J'aime à **sentir**, sous mes pieds, la butte de pierre et de terre à la gloire du dieu, pleine et dure : pas de creux, ...
5. Cela donne à la pièce un air de fête, de désordre joyeux où l'on pourrait presque **se sentir** bien, ...
6. ... et ma venue au monde m'a traumatisé à jamais, si bien que, depuis, je ne cesse de dépérir et de **me sentir** nu, comme un arbre à qui on aurait, d'un seul tour de main, enlevé son écorce.
7. nous couvrons de crème aux agrumes pour **sentir** le pamplemousse: cette odeur nous donne toujours un air bronzé.

Non-reflexive uses	Reflexive uses
<i>le pamplemousse</i>	<i>trop seule</i>
...	...

What generalizations can you draw about the types of complement (noun, adjective, or adverb) which follow the non-reflexive and reflexive uses? Write two original sentences illustrating this difference.

Example 3. Phrasal verb: (*locution verbale*): *mettre en* + noun

Task: discern the meaning of one or more phrasal verbs from a set of Key-Word-In-Context (KWIC) lines.

Procedure: divide the class into small groups, give each group a small set of concordance lines for two or three phrasal verbs (*mettre en œuvre*, *en doute*, *en garde*, *en place*, *en pratique*, *en péril*, *en valeur*, *en évidence*, *en lumière*). The group must propose an appropriate translation or paraphrase for each expression, note its typical complements, and write an original sentence for each. Groups present their findings to the class. Follow-up activity: gap-filling exercise based on concordance lines for all expressions studied.

Sample concordance source: Sharoff, Leeds Internet Corpus (see section 5 below).

1. qui se posent dans les salles d'urgence. Nous allons aussi **mettre en oeuvre** la Stratégie ontarienne de prévention des accidents
2. Coût: le coût comprend le coût réel actuel pour **mettre en oeuvre** la mesure d'adaptation ainsi que tous les coûts
3. " Nous sommes déterminés à **mettre en oeuvre** la prochaine génération de mesures qui assurera
4. endroit au bon moment, avec l'équipement nécessaire pour **mettre en oeuvre** le plan élaboré par l'officier.
5. un incident grave, et à déterminer le moment et la façon de **mettre en oeuvre** le programme.

Example 4. Multi-word unit: the phrasal conjunction *de sorte que*

Task: match the initial main clause and the appropriate subsequent dependent clause introduced by *de sorte que*.

Instructions to student: trouvez la bonne terminaison de chaque phrase.

Concordance source: Lexiquum (La Presse).

1. Ces modèles nous ont donné une très bonne idée de l'endroit où allait frapper Hugo,
2. Ce que les médecins appellent l'obésité abdominale est un facteur de risque très important
3. Les assaillantes ont pris la fuite avant l'arrivée de la police,
4. Tout ce temps, Toronto prenait les devants en matière de design et de bureaux de ventes;
5. Nous savions tous qu'il y avait eu une enquête en 1975 et qu'elle n'avait rien donné,
6. ... ont toujours eu la faveur des écoliers. Si vous prévoyez une orange, entaillez-la au préalable

-
- a. ... **de sorte que** personne n'a été arrêté relativement à cette affaire.
 - b. ... **de sorte que** j'ai pensé qu'il était absolument inutile d'aller à la police quelques années plus tard.
 - c. ... **de sorte que** l'industrie montréalaise de la mode se trouve en situation périlleuse...
 - d. ... **de sorte que** l'enfant puisse la peler facilement.
 - e. ... **de sorte que** les gens qui ont un tablier c'est-à-dire une bedaine pendante sont plus sujets aux crises cardiaques
 - f. ... **de sorte que** les autorités ont pu intervenir à temps et prendre les mesures d'évacuation qui s'imposaient,

Example 5. Discourse function of *il s'agit de ...*

Task: describe the discourse function of *il s'agit de ...* on the basis of several examples in extended contexts.

Sample concordance: various websites

Instructions to student: study the following examples of the use of the verb *s'agir (de)* and answer these questions:

1. What do you notice about the subject of this verb? To what does this subject refer?

2. Try to describe what the expression *il s'agit de* is doing in each context. What other French expression could replace it? What is the relationship between what precedes the expression and what follows it? Do you think there is any significance to the fact that all three examples involve direct quotes from an apparent 'expert'?
1. "Je suis très surpris. Je crois que c'est là la réaction d'une personne qui n'a pas résisté au stress. **Il s'agit d'**un employé qui disjoncte", avait répondu M. Nicolaï, ...
 2. WASHINGTON (AFP) - Des scientifiques ont utilisé en laboratoire la thérapie génique pour aider des cellules du coeur endommagées à fonctionner de nouveau normalement, selon une étude publiée lundi dans la revue de l'American Heart Association. **Il s'agit d'**"une étape vers d'éventuels essais cliniques sur des humains", a déclaré l'équipe dirigée par le Dr Walter Koch, du Jefferson Medical College, à l'Université Thomas Jefferson de Philadelphie (Pennsylvanie, nord-est).
 3. MONTREAL (AP) - Un incendie d'origine criminelle a détruit dans la nuit de dimanche à lundi la bibliothèque de l'école primaire juive United Talmud Torah à Montréal: pour la police **il s'agit d'**un "crime à caractère haineux" et antisémite qui semble être revendiqué par une organisation inconnue.

Discussion: a small number of occurrences in their extended context provide the basis for reflection on the function of the expression with respect to the preceding discourse. Students can be led to notice, in addition to the non-referential nature of *il*, that what is introduced by *il s'agit de* is a new description of an event described in the preceding discourse, and that this description adds some new information which usually entails some evaluative judgment of the event or situation.

Example 6. Syntactic contrast: *il est* + Adj + *de* + Inf vs. NP *est* + Adj + *à* + Inf
Task: distinguish two similar syntactic patterns that frequently occur with *facile* / *difficile* and supply the appropriate preposition before the infinitival complement.

Discussion: this example involves study of the word grammar of *facile* and *difficile* with infinitival complements, the one construction involving non-referential *il* and the preposition *de*, the other a referential subject and the preposition *à*. This pair of confusable constructions is of course generalizable to other adjectives that occur with a non-referential subject pronoun (*possible*, *impossible*, *bon*). The sample activity is a gap-filling verification exercise which would be preceded by an activity involving study of a similar concordance without gaps, where learners would be led to observe the differences between the two syntactic patterns and their semantic correlates.

Quelle préposition faut-il? *à* ou *de*?

Concordance source: Lexiquum (various subcorpora).

1. Il a saisi l'image. Notre premier sujet de conversation est donc facile ____ trouver : l'absence de Maria, son père et tout ce qui s'ensuit. Il se détourne un peu
2. Je compte sur elle pour prévenir les femmes de l'état de santé de Tibé. La vérité est plus facile ____ entendre dans le recueillement d'un cloître et celui des révérendes
3. Il est toujours plus facile ____ suivre des grands idéaux que d'accepter d'être un peu ridicule.
4. le climat a un peu la douceur de celui de la Floride. J'espère y amener les parents. Ça ne sera pas facile ____ persuader papa qui, depuis son accident à la jambe, a perdu l'habitude des voyages.
5. selon un rapport de la DEA publié cet été. Ce puissant dérivé de la cocaïne, facile ____ se procurer et d'un prix modique - une dizaine de dollars la dose - est actuellement disponible

Example 7. Translation synonyms *en* and *dans*.

Task: examine a somewhat lengthy concordance of each preposition to observe a syntactic difference in their objects, and to note common collocations and differences in collocating behavior.

Discussion: while an edited concordance prepared and printed by the teacher can be used for this purpose, a more effective approach would be to have students work with one of the user-friendly online concordancers (see section 5), or at least a lengthy print-out for each preposition. An initial comparison of the two concordances will reveal an important syntactic generalization: while *dans* is always followed by a determiner, *en* is almost never followed by one. Next, students can be asked to look for common collocations of each (assuming the concordancer allows sorting according to the right-hand context) and to indicate, if possible, with what kinds of words each one commonly occurs. At lower levels, this will lead to noticing the use of *en* with months, years, and certain geographical names. In addition, learners can be asked, after some familiarization with the concept of set phrases, to determine if there are any set expressions among the collocations. This question should lead to the observation that *en* occurs in very many fixed expressions, whereas *dans* occurs in relatively fewer such expressions. Finally, learners should be led to observe that *en* in fact represents two different lexical items, the pronoun and the preposition. In turn, within the category of prepositional uses, the use of *en* with the present participle to form the gerund can be distinguished from the other more numerous cases.

5. Comparison of online corpora / concordancers

Because websites that combine corpora and easy-to-use concordancers offer teachers the easiest approach to creating their own concordance-based learning materials, I will describe here what I have found to be the most useful among such sites that are not password-protected. The table in Fig. 1 (see below) compares the features of the following four sites:

- Compleat Lexical Tutor:
http://www.lextutor.ca/concordancers/concord_f.html (17.05.2013)
- Lexiquum: <http://atour.iro.umontreal.ca/cgi-bin/lexiquum> (17.05.2013)
- CorpusEye: <http://corp.hum.sdu.dk/cqp.fr.html> (17.05.2013)
- Leeds Internet Corpora: <http://corpus.leeds.ac.uk/internet.html> (17.05.2013)

Designed by Tom Cobb for language teachers and their students, the Compleat Lexical Tutor (CLT) is the most user-friendly of the four sites, as shown by its use of ordinary English terms in the interface, as in the choice of 'Equals', 'Starts', 'Ends', or 'Contains' to search for variant forms of a lexical item or all lexical items with certain endings or base forms⁶. Also helpful is the ease of exporting selected lines to a Word document which even preserves the KWIC formatting. However, the small size of the CLT corpus is limiting, especially if one prefers to use more contemporary texts. In selecting corpora, one should keep in mind that learners' success at interpreting concordance lines will be facilitated when the corpus searched consists of text-types and kinds of topics which are familiar to the learners.

The Lexiquum concordancer is also very user-friendly, though it offers the least versatility, lacking both the capability of sorting KWIC lines by words in the immediate context of the keyword, and the provision of any information about collocations. However, desirable features are the large size of the corpus and the variety of text-types offered. I have found the literary corpus (Editions Leméac: novels, plays and poetry by 15 authors, published 1991-1993) to be very useful due to the relative accessibility of its language and its inclusion of personal narratives and dialogue.

⁶

It should be noted that this site contains many other software programs for researchers, teachers and learners of French and English, such as vocabulary profilers and hypertext and cloze builders. See Sevier (2004) for a review of the site.

	Compleat Lexical Tutor	Lexiquum	CorpusEye	Leeds Internet Corpora
CORPORA				
<i>Text types</i>	Le monde 1998 (1.1 M words), 3 19th-c. authors, 2 nonfiction	Largest: journalistic, leg. debates & commissions; small lit. & scient.	Parliamentary debates (29 M), Wikipedia 2005 (37.8 M)	General- purpose Internet corpus (open source development)
<i>Size (total)</i>	2.8 M words	229.3 M words	66.8 M words	200 M words
OUTPUT				
<i>No. of hits</i>	Variable, < 5,000	Variable, < 500 per corpus	Unlimited	Variable, < 1,000
<i>Sort options</i>	1-3 words L / R	None	None	1 word L / R
<i>KWIC line length</i>	Variable, <140 char.	Variable, <200 char.	Fixed, <100 char.	Variable
<i>Extended context</i>	Paragraph-length	Leg. debates & commissions only	Varies, 1-3 sentences	Yes, variable length
<i>Identified source</i>	No	Yes	Wikipedia only	Inconsistent, URL
VERSATILITY				
<i>Phrasal search</i>	Yes	Yes	Yes	Yes
<i>Variant forms</i>	'Equals / starts / ends / contains'	Add '+' to end of lemma	Reg. express. (.* = wildcard)	Reg. express. (.* = wildcard)
<i>Discontinuous forms</i>	No	Yes ('.. = up to 25 characters)	Yes: 'refined' or CQP ⁷ search	Yes: CQP search
<i>Collocations</i>	Yes	No	Yes, multi- option menu	Yes, multi- option menu
<i>Export lines</i>	Very easy: all, selected, or random, in 1 click	Manual only	Single lines or all, with 1 click	Manual only

Fig. 1. Comparison of selected online concordancers

Both CorpusEye (CE) and the Leeds Internet French Corpus (LIC) introduce a greater degree of complexity in their interfaces, due to their greater versatility. Namely, they both offer the option to include syntactic indicators in the search expression. LIC provides a simple means of adding syntactic labels to lexical search terms, which can be helpful, for instance, in eliminating homonymous forms of a different syntactic category. Both CE and LIC offer very sophisticated search options through the use of regular expressions and the special CQP-search language, but both concordancers can also be used for ordinary lexical searches without the use of these more advanced tools. CE's

⁷ Corpus Query Processor.

sizable Wikipedia 2005 corpus offers a text-type that is certainly familiar to most learners and covers a wide variety of topics. LIC presents considerable flexibility and ease of use, for instance in its multiple sorting options, but it also has certain disadvantages. For instance, though the concordancer can be set to search for up to 1,000 hits, the contexts generated are always the first ones encountered in the corpus, rather than a sampling throughout, which can result in a large proportion of occurrences from the same source. Both CE and LIC offer extensive information about frequencies of collocations, and LIC includes a helpful option to access concordance lines corresponding to each of the collocations listed.

Though I have described above what I believe to be the most readily usable concordancing sites, there are of course other online databases that allow similar searches. For instance, the FRANTEXT database (<http://www.frantext.fr/>, 17.05.2013), a collection of almost 3,000 literary works, predominantly of the last three centuries, now offers a very user-friendly interface with multiple search capabilities. Though conceived and developed primarily for literary and linguistic research, FRANTEXT is quite usable for pedagogical purposes, as it allows the limiting of searches by author, title, period and / or genre. Output is variable from 500 to 10,000 occurrences, with various sorting options and collocation tables. A possible drawback, however, is the limitation of access to institutional subscribers. Also of interest is the Leipzig University's Wortschatz site that includes two search engines of interest. The first (<http://corpora.informatik.uni-leipzig.de/>, 17.05.2013) offers searches of 230 corpus-based monolingual dictionaries and provides information on frequency and collocations, as well as an extensive list of complete-sentence examples with URLs of their source sites. The second (http://wortschatz.uni-leipzig.de/ws_fra/index.php, 17.05.2013) searches a corpus of 700 million words drawn from Francophone newspapers, websites and Wikipedia articles (see Kamber 2011). A disadvantage of both is that the examples are not presented in KWIC format, so they occur in no particular order, making it impossible to abstract patterns of occurrence from the data. Also, multi-word searches are not possible.

All of the foregoing are sites whose corpora consist of written texts. Online corpora of spoken French are still considerably less numerous than the latter. There is however the SACODEYL site (<http://www.um.es/sacodeyl/>, 17.05.2013), a small pedagogic corpus focusing on European youth language, which includes searchable transcriptions of about 4 hours of interviews with French young people, as well access to the video recordings. The ELICOP site (<http://bach.arts.kuleuven.be/elicop/>, 17.05.2013) offers searchable access to several large spoken corpora, but it lends itself better to research than to pedagogical purposes, due to the linguistic nature of the transcriptions.

Of course, ordinary web browsers may be used as concordancers and are quite usable for a quick check of spelling or grammatical construction. However, for more sophisticated questions of usage or collocations, the output of such search engines is often problematic in both content and format. For this reason, computational linguists have devised software to improve upon the output of web searches (Kilgarrieff & Grefenstette 2003), one example of which is the online WebCorp concordancer (<http://www.webcorp.org.uk/live/>, 17.05.2013).

Finally, there is also the possibility to download corpora such as the Chambers-Rostand Corpus of Journalistic French and the Chambers-Le Baron Corpus of Research Articles in French, both available from the Oxford Text Archive (<http://ota.ahds.ac.uk/>, 17.05.2013). Corpora such as these can be searched using software such as AntConc, a fairly user-friendly program available at Laurence Anthony's AntLab site (http://www.antlab.sci.waseda.ac.jp/antconc_index.html, 17.05.2013).

6. Conclusion

As suggested above, concordance-based activities can facilitate the acquisition of both lexis and grammatical forms. With respect to grammatical items, observation of data and hypothesis formulation can help to focus on the role of grammatical forms in producing differences of meaning, thus conveying the fundamental principle that communicating effectively is not primarily a matter of following rules but of making appropriate choices corresponding to alternative meanings. Shelly (1993) suggests a similar kind of approach to complex grammatical distinctions; the kind of teacher-guided, meaning-focused inductive dialogue modeled in her article would be perfectly complemented by a concordance-based observation activity.

Some proponents of DDL have pointed to the fact that a teacher's implementation of DDL requires a certain acceptance on the teacher's part of a reduced degree of control. Gilquin & Granger (2010: 366-367) describe it this way:

Although [...] DDL activities may be situated along a continuum from teacher-led to learner-led, it is nonetheless true that teachers have a less central role in DDL than in traditional teaching. They tend to have relatively little control over what happens during the lesson, which may be 'incompatible with the "minimum risk" scenario which can be found in many teaching cultures' (Boulton 2009: 93). In addition, since the computer becomes the main source of knowledge, this may be experienced as a 'loss of expertise' by the teacher (Hunston 2002: 171). DDL, therefore, requires that teachers take risks, and agree to 'let go' and let the student take pride of place in the classroom.

Though the kinds of activities I have described clearly fall toward the teacher-led end of the continuum, as students progress in their use of DDL, this aspect does need to be recognized by the teacher. At the same time, however, the use of concordance-based activities does not lessen the teacher's workload

when it comes to planning and creating activities. In the words of Lewis (1993: 149):

Far from being a soft option for teachers, this imposes considerable demands on them. Their skill lies in perceiving actual or potential confusions in their student's perception of a system of English [or other target language], and devising a task accompanied by language data focusing on the particular language problem. Teachers must choose when to intervene and provide appropriate tasks, questions and data. That is difficult enough, but then they must resist the temptation to interfere, and allow students the time to discuss, formulate, and if necessary reformulate their perceptions.

While the factor of the extra time required for DDL activities will have to be reckoned with by the teacher, if we are serious about creating effective learning conditions, we should not be satisfied with reliance on 'imposed' explanation. It is therefore my hope that more second language educators will accept the challenge to explore the potential benefits of data-driven learning.

REFERENCES

- Aston, G. (2001). Learning with corpora: an overview. In: G. Aston (ed.), *Learning with corpora* (pp. 7-45). Houston: Athelstan.
- Boulton, A. (2011). Data-driven learning: the perpetual enigma. In: S. Gozdz-Roszkowski (ed.), *Explorations across languages and corpora* (pp. 563-580). Frankfurt: Peter Lang.
- Chambers, A. (2010). What is data-driven learning? In: A. O'Keeffe & M. McCarthy (eds.), *Routledge handbook of corpus linguistics* (pp. 345-358). London: Routledge.
- Cheng, W. (2010). What can a corpus tell us about language teaching? In: A. O'Keeffe & M. McCarthy (eds.), *Routledge handbook of corpus linguistics* (pp. 319-332). London: Routledge.
- Cobb, T. (1999). Applying constructivism: a test for the learner as scientist. *Educational Technology Research & Development*, 47, 15-31.
- Collins COBUILD (2006). *Advanced learner's English dictionary*. New York: HarperCollins.
- Conklin, K. & Schmitt, N. (2008). Formulaic sequences: are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29, 1, 72-89.
- Gabrielatos, C. (2005). Corpora and language teaching: just a fling or wedding bells? *Teaching English as a Second Language – Electronic Journal*, 8, 4, 1-35. <http://tesl-ej.org/ej32/a1.html> (15.05.2013).
- Gilquin, G. & Granger, S. (2010). How can data-driven learning be used in language teaching? In: A. O'Keeffe & M. McCarthy (eds.), *Routledge handbook of corpus linguistics* (pp. 359-370). London: Routledge.
- Hulstijn, J., Hollander, M. & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: the influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80, 327-339.
- Hunston, S., & Francis, G. (2000). *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Philadelphia: Benjamins.
- Johns, T. ([1990] 1991). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, 10, 14-34. [Revised version: T. Johns & P. King (eds.), *Classroom concordancing*. *English Language Research Journal*, 4, 27-45].
- Johns, T. (1993). Data-driven learning: an update. *TELL&CALL*, 2, 4-10.

- Kamber, A. (2011). Contexte et sens: utilisation d'un corpus écrit dans l'enseignement/apprentissage du FLE. *Travaux neuchâtelois de linguistique*, 55, 199-218.
- Kilgarriff, A., & Grefenstette, G. (eds.) (2003). Special Issue on web as Corpus. *Computational Linguistics*, 29.
- Lewis, M. (1993). *The lexical approach: the state of ELT and the way forward*. Hove, England: Language Teaching Publications.
- Lewis, M. (1997). *Implementing the lexical approach: putting theory into practice*. Hove, England: Language Teaching Publications.
- Moudraia, O. (2001). *Lexical approach to second language learning*. Washington, D.C.: ERIC Clearinghouse on Language and Linguistics, Center for Applied Linguistics. EDO-FL-01-02. <http://www.cal.org/resources/digest/0102lexical.html> (09.08.2007 from CAL database).
- Nattinger, J. & DeCarrico, J. (1989). Lexical phrases, speech acts and teaching conversation. In: P. Nation & R. Carter (eds.), *Vocabulary acquisition*. *AILA Review*, 6, 118-39.
- Nattinger, J. & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: the role of corpus consultation literacy. *ReCALL*, 19, 3, 269-286.
- Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: native-like selection and native-like fluency. In: J. Richards & R. Schmidt (eds.), *Language and communication* (pp. 191-226). London: Longman.
- Peters, A. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: a case study. *Reading in a Foreign Language*, 18, 1-28.
- Römer, U. (2006). Pedagogical applications of corpora: some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik*, 54, 121-134.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Sevier, M. (2004). Review of the compleat lexical tutor. *Teaching English as a Second Language – Electronic Journal*, 8, 3, 1-10. <http://tesl-ej.org/ej31/m2.html> (15.05.2013)
- Shelly, S. (1993). Rule reformulation at the advanced level. *French Review*, 66, 760-768.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2000). Lexical grammar. *Naujoji metodologija*, 24, 191-202. <http://donelaitis.vdu.lt/publikacijos/sinclair.pdf> (09.23.2007)
- Sockett, G. (2011). Autonomie de l'apprenant et environnement personnel d'apprentissage. *Les Langues modernes*, 105, 22-31.
- Stevens, V. (1991). Concordance-based vocabulary exercises: a viable alternative to gap-fillers. In: T. Johns & P. King (eds.), *Classroom concordancing*. *English Language Research Journal*, 4, 47-63.
- Willis, J., & Willis, D. (1989). *Collins COBUILD English course*. London: Collins COBUILD.

