AAUSC 2008 Volume

Conceptions of L2 Grammar: Theoretical Approaches and their Application in the L2 Classroom

Stacey L. Katz and Johanna Watzinger-Tharp

HEINLE
CENGAGE Learning™

AAUSC 2008 Volume, Conceptions of L2 Grammar: Theoretical Approaches and their Application in the L2 Classroom
Stacey L. Katz
Johanna Watzinger-Tharp

Executive Editor: Lara Semones

Assistant Editor: Cat Thomson

Editorial Assistant: Cat Mooney

Marketing Manager: Lindsey Richardson

Marketing Coordinator: Jill D'Urso

Senior Art Director: Cate Barr

Print Buyer: Susan Carroll

Production Service/Compositor: Pre-Press PMG

For product information and
technology assistance, contact us at Cengage Learning
Customer & Sales Support, 1-800-354-9706

For permission to use material from this text or product, submit all requests online at **cengage.com/permissions**
Further permissions questions can be emailed to
permissionrequest@cengage.com

Cengage Learning
25 Thomson Place
Boston, MA 02210
USA

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

For your course and learning solutions, visit
academic.cengage.com

Purchase any of our products at your local college store or at our preferred online store **www.ichapters.com**

Chapter 8

Applications of Corpus-Based Linguistics to Second Language Instruction: Lexical Grammar and Data-Driven Learning

Betsy J. Kerr, University of Minnesota

Conrad (2000) poses a key question: Will corpus linguistics revolutionize grammar teaching in the twenty-first century? Addressing English as second language (ESL) professionals, Conrad notes that the concurrence of a rebirth of interest in grammar instruction and the proliferation of corpus-based grammar studies during the latter years of the twentieth century did not lead to a coming together of these two strands in a way that would significantly affect grammar instruction in the ESL classroom (pp. 548–549). The absence of such an effect is even more evident in the field of foreign language (FL) education in this country, as very few FL practitioners at any educational level are aware of the potential applications of corpus-based research and its methodology in FL education. Conrad suggests the following three changes as potential benefits for ESL instruction:

1. Replacement of monolithic grammatical descriptions with register-specific descriptions
2. Greater integration of vocabulary and grammar instruction
3. Greater emphasis on appropriate conditions of use for alternative grammatical constructions (p. 549)

My own call to FL professionals—and to FL program directors in particular—has slightly different emphases, as presented in this chapter. First, language program directors need to become aware of the potential role of corpus-based research in shaping the grammatical syllabus at any given level. Second, I strongly echo Conrad's appeal for greater integration of lexical and grammatical content in the FL curriculum, drawing support from a conception of grammar derived from corpus-based research and supported by other branches of inquiry as well. Finally, focus-on-form instruction should be enriched through the use of consciousness-raising tasks that engage learners in discovery procedures similar to those of corpus linguistics or that employ teacher-adapted concordancer output.

Some Preliminaries: Introduction to Corpora and Concordancers

A corpus is "a large and principled collection of natural texts" on which a linguistic analysis is based (Biber, Conrad, & Reppen, 1998, p. 12). When employed with a concordancer, a corpus is assumed to be an electronically formatted text or set

of texts. A corpus is principled in that the texts are selected to be representative of a particular type of real-world language use, whether written or oral, with the latter having been transcribed from audio recordings.

Corpora are analyzed by means of specialized computer software, known as concordancers, which generate for a given word or expression a list of all occurrences in the given corpus, with a variable amount of the text surrounding each occurrence. The standard format for concordancer output is the key word in context (KWIC) format, illustrated by the following concordance of Fr. *incluant* ("including") generated by the online French concordancer that is part of Cobb's (2007) *Compleat Lexical Tutor:*

Concordances for *incluant* = 5

1. mars, 500 victimes de mutilations, *incluant* des amputations, avaient été admis
2. s secrets de toutes sortes d'armes *incluant* des missiles anti-tanks type SAM-7
3. e (CAP): 300 millions de francs, *incluant* la restauration du bâtiment, la ré
4. de francs d'économies, mais en y *incluant* le récent accord passé avec la CNAM
5. cher les 50 millions de francs en *incluant* les résultats du Monde diplomatique

Concordancer Features

Over the last decade, online concordancers have become so readily available and so user-friendly that concordancing is now readily accessible to all educators of commonly taught languages, for their own use as well as for the use of their students. Of course, online concordancers vary considerably with respect to a number of features, of which the following are of potential interest:

1. Choice of corpora or combinations of corpora. These may vary according to medium and text type; sites differ considerably in the degree of variety, with some specializing in one type and others presenting a range of genres. Corpus size varies considerably.
2. Choice of the length of the context provided, and the related possibility of consulting a significantly larger context of individual examples.
3. Choice of the number of occurrences obtained.
4. Nature of the search expression. Most concordancers allow searches for multiple-word expressions, and many provide methods of searching for variant forms of a lexical item. The most sophisticated interfaces allow the use of "regular expressions" to locate not only variant forms of a lexical item, but also one or more items that co-occur but are not contiguous. Some interfaces provide their own simplified version of regular expressions, thereby increasing their versatility while remaining user friendly.
5. Choice of sorting order (order in which search results appear). The default option is generally the order of occurrence in the text. Other very

useful options are "sort right" (alphabetical order based on the word that directly follows the keyword) and "sort left" (alphabetical order according to the word that immediately precedes the keyword). These options are essential to the identification of collocations and lexical phrases.

6. Possibility of searches employing syntactic information. Such searches are possible only where the corpus has been "tagged"—that is, parsed and annotated according to syntactic categories. Few online concordancers currently offer this option.

7. Possibility of displaying specific source information.

8. Format of search results. Though the KWIC format illustrated in Example 1 (later in this chapter) is the standard, some variations are to be found.

The appendix to this chapter provides further information about specific online concordancers for the commonly taught languages. Of course, one can also use any web browser for the same purpose, though the abundance of redundant or extraneous items in the output often requires considerable manual filtering by the user. If one wished to study a common grammatical form, one could, of course, use any online text of a certain length and employ the word processor's Find function as a primitive concordancer.[1] The other option that must be mentioned is the use of concordancing software, of which several low-cost or free versions are available; these programs allow one to perform concordancing with any electronic text or, in some cases, with webpages of the user's choice.[2]

Corpus Linguistics

The development of electronic text databases and the means to search such texts gave rise, starting in the 1960s, to the field known as corpus linguistics or corpus-based linguistics. The essential principle of corpus-based investigation is that "it is empirical, analyzing the actual patterns of use in natural texts," as opposed to seeking to describe what is theoretically possible in a language (Biber et al., 1998, p. 4). Corpus linguistics seeks to discover the patterns that characterize language use—that is, the frequently occurring associations of language features, where the features in question may be lexical items or grammatical structures. Patterns studied may also involve the association of linguistic features with nonlinguistic features such as register, dialect, or time period.

Corpus-Based Research and FL Curricula

One application of which language program directors should be aware is the potential role of corpus-based research in shaping the grammatical content of language curricula. By providing both quantitative and qualitative information, such as frequencies of occurrence and functional descriptions of particular grammatical forms in a given corpus, corpus-based studies can lead to more refined accounts of native-speaker uses of the grammatical forms in question. Such

accounts can offer a corrective to traditional pedagogical descriptions that may rely too heavily on native-speaker intuitions or, often, on characterizations of formal written language, when the linguistic register (or registers) set as the curricular goal may be less formal or more diversified.

There exists an abundance of corpus-based studies treating various aspects of English grammar. These materials have already been translated into pedagogical resources such as the *Collins–Cobuild English Grammar* (1990) and the *Longman Grammar of Spoken and Written English* (Biber, Johansson, Leech, & Conrad,1999), both of which are based on corpora of British and American English. Unfortunately, the field has seen much less work on the grammars of other languages. Nevertheless, there are some significant studies in these areas, and their numbers appear to be increasing. For Spanish, there is the recent comprehensive work by Biber et al. (2006), as well as more limited studies such as those carried out by Stenström (2006), Ramón García (2002), and Sancho Sánchez (1999). For French, Di Vito's (1997) study remains the most extensive work, but many other studies treat particular aspects of the grammar: Recent references include Anderson (2007) for adjective placement, Polzin-Haumann (2003) on the relative pronoun *dont* ("*of which, whose*"), Marnette (2003) on indirect discourse, Fonseca-Graber & Waugh (2003) on subject pronouns, and Labeau (2006) for past tenses. Studies of German tend to focus on modal particles and prepositions; representative titles include the studies by Möllering (2004), Lutzeier (1999), Belz (2005), and Dodd (2006).

Language program directors should be aware of these advances in grammatical description and their implications for evaluation and selection of course texts, as well as in relation to their frequent role as textbook authors. While many factors must be taken into consideration in crafting a grammatical syllabus (Anderson [2007] offers an insightful examination of the place of corpus-based research findings among these considerations), program directors should look for evidence of research-based considerations in the grammatical content of potential texts, especially at intermediate and advanced levels of instruction. Directors should also be on the lookout for new pedagogical materials that make use of corpus linguistics methodologies, such as Zinggeler's 2007 publication, *GRIMMATIK: German Grammar through the Magic of the Brothers Grimm Fairy Tales* (see Zinggeler, 2006, 2007).

Implications for our Conception of Grammar: The Lexis–Grammar Relationship

Sinclair's Lexical Grammar

John Sinclair, the driving force behind the creation of the massive Cobuild corpus of English, was a major figure in the development of corpus linguistics in all its aspects—theoretical, descriptive, and applied. The focus of much of Sinclair's work was collocations, or groups of lexical items that "commonly occur together and therefore, for the language user, constitute a single language choice rather

than a series of choices" (Hunston & Francis, 2000, pp. 230–231). The originality in Sinclair's approach to collocations lay in his extension of the notion of phraseology to cover a much greater portion of language than had been previously proposed; for Sinclair, these "semi-preconstructed" phrases—more or less predictable, but not fixed—are the general rule in language, rather than the exception.

Because this notion of semi-preconstructed phrases or "chunks" of language as meaningful units is so central to Sinclair's conception of grammar and the work built on it, it is illustrated here with an example drawn from Sinclair as reported by Hunston and Francis (2000)[3]—namely, the phrase *naked eye*. Inspection of even a short concordance[4] 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 . . .); 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* (Hunston & Francis, 2000, pp. 24–25).

Sinclair concludes from these observations that *naked eye* is part of a larger "meaning unit" that 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": These semi-preconstructed phrases actually represent a single choice made by the language user, despite their appearing to be analyzable into segments (Sinclair, 1991, p. 110, cited in Hunston & Francis, 2000, p. 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, p. 195). Evidence from corpora of the prevalence of such language chunks led Sinclair to call into question the very distinction between lexis and grammar and to propose the notion of a "lexical grammar" where the grammar and lexis are built together, with each on an equal basis (p. 191).

As Hunston and Francis (2000, pp. 11–12) point out, Sinclair's work is paralleled by research in related fields that places in the foreground the role of lexical phrases—that is, relatively fixed sequences of morphemes, designated with various terms such as "gambits," "sentence stems," and "formulae." Although Sinclair takes a more extreme position, suggesting that *all* language is patterned, these other currents of research nonetheless converge in a greater recognition of the phraseological nature of language. In the field of language teaching, authors such as Pawley and Syder (1983) and Nattinger and DeCarrico (1989, 1992) highlight

the role of the acquisition of lexical phrases in the achievement of fluency and native-like language production. In psycholinguistics, lexical phrases have also figured prominently in debates about L1 acquisition and the way in which language is encoded in the brain, whether as single lexical items to be combined with the help of grammatical rules or as multiword 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 structures.

Porto (1998) argues that the preceding findings would not be sufficient in themselves to justify the choice of the lexical phrase as an ideal unit for teaching. She goes on to justify this choice with the following line of reasoning, based on the aforementioned research: Because lexical phrases are stored in the lexicon as unanalyzed chunks, they are easily retrievable by learners and allow learners to develop conversational fluency at early stages without the need to analyze the structure of these phrases. However, lexical phrases have a dual nature, because they are analyzable by the rules of grammar and thus serve as a springboard for language development; a simple lexical phrase associated with a particular functional use can later be expanded to express the same function in increasingly more difficult ways. Moreover, lexical phrases are easy to acquire due to their high frequency and their recurrent association with a particular situational context, both of which make them highly memorable for learners.

Following Sinclair's conception of a lexical grammar, Conrad (2000, pp. 552–553) argues for a better integration of grammar and vocabulary in pedagogical materials and activities to do justice to the complex connections between grammar and the lexicon—namely, the fact that particular grammatical constructions or forms tend to occur more frequently with particular lexical items. Conrad's illustration is that of the association of English *that-* and *to*-complement clauses with differing groups of matrix verbs (*think, say, know, see* versus *want, try, like*). Although Conrad acknowledges that grammars often furnish lists of such verbs, these lists are often not consistent with actual usage. My own assessment of FL textbooks is that grammar lessons are rarely accompanied by useful lexical sets, and even when they are, the lexical material is not sufficiently exploited in the learning activities.

Pedagogical Applications: Lewis's Lexical Approach

One of the most significant attempts to apply the insights of the foregoing research to second language pedagogy is Lewis's lexical approach (Lewis, 1993, 1997). Lewis's principles lend themselves well to the American FL context, as they are couched in terms of modifications that can be made to conventional communicative or grammar-based teaching practice (as opposed to a whole new model that would replace the existing practice). This section highlights some of Lewis's basic principles and important methodological recommendations, especially as they pertain to grammar teaching and the use of corpora and concordancing.

Although Lewis acknowledges the usefulness of structural patterns, given the fallacy of the grammar/vocabulary dichotomy, he would allot a much greater proportion of instructional time and effort than is currently practiced to explicit instruction in lexical patterning. Preferring the term *lexis* to *vocabulary* due to the

latter's association with individual words, the emphasis in the lexical approach is on multiword units, of which Lewis enumerates several types (1993, pp. 8–11):

- Collocations (words that co-occur in natural text with greater than random frequency), such as *miss the bus, make a mistake,* and *slump dramatically.*
- Fixed expressions, which include social greetings, politeness phrases, and idioms, but also many more "prefabricated multiword items" such as *How long will it take? What size do you take?* and *I can't take any more!,* to cite just a few examples of the myriad uses of the de-lexicalized verb *take.*
- Semi-fixed expressions, "a large and important category which contains a spectrum, from very short to very long and from almost fixed to very free" (p. 11). These include expressions with a "slot" such as *Could you pass . . . , please?* and *Nice to see you. I haven't seen you* + time expression with *for* or *since,* and "sentence heads" such as *What was really interesting / surprising / annoying was . . . .*

A central element of Lewis's approach has to do with raising learners' awareness of the phrases or "chunks" of which utterances are composed, so many activity types have as their goal the development of the ability to "chunk" language successfully (i.e., to parse it into sense groups). Such activities include first and second language comparisons and translation—albeit carried out chunk-for-chunk, rather than word-for-word.[5] Other activities focus on guessing the meaning of vocabulary items from context or identifying collocations in texts. To keep active the new words and expressions that have been learned, Lewis urges the recycling of activities, such as summarizing a text orally one day and then doing so again a few days later. While doing intensive and extensive listening and reading in the L2, Lewis recommends that learners be trained to create a lexical notebook where, to quote Woolard (2000, p. 43), they "record, revisit and re-activate the significant vocabulary they meet." Working with dictionaries and other reference tools, especially electronic ones, is also encouraged, but to explore a word's collocational field rather than simply to answer the question "What does X mean?" (Woolard, 2000, p. 36). Many detailed descriptions and examples of lexis-focused activities can be found in Lewis (1997).

Despite Lewis's (1993) de-emphasis of grammar in favor of lexis, he does acknowledge a place in the language syllabus for grammar, when rightly understood. Seeing grammar as "the search for powerful patterns" (p. 137), Lewis suggests increased attention to word grammar, the set of patterns in which a word typically occurs, and de-lexicalized words. Pedagogical strategies for the study of word grammar include the following: identifying and distinguishing among high-frequency uses; identifying and recording high-frequency collocations; and identifying derived or morphologically related forms and their collocations (p. 142). De-lexicalized words have low meaning-content, such as the English verbs *make, have,* and *take* and the function words *with, by,* and *on*. *With*, for example, generates patterns where it co-occurs with words denoting a physical feature, a tool, a person or people, a mood, an illness, or an institution (p. 144).

In his discussion of the "how-to" of grammar instruction, Lewis downplays the two pillars of conventional language pedagogy, explanation and practice, noting that "the essence of language is meaning, and meaning implies choice," which is necessarily restricted by controlled practice (1993, p. 148). In place of explanation, he advocates engaging the learner in exploration. In fact, he recommends that both the teacher and the student adopt the critical attitude and scientific processes of the linguist, engaging with real language data through the observe–hypothesize–experiment procedure. Accordingly, Lewis suggests, much of grammar work will be receptive rather than productive, and student centered rather than teacher centered. Receptive activities will be directed toward awareness-raising, asking students to notice certain grammatical features of a text that has been used for another activity, or requiring students to process items through activities such as checking, sorting, comparing, and matching. Finally, in this perspective, productive grammar practice should be seen as hypothesis-testing experimentation, where the primary goal is *successful* language rather than accurate language.

Classroom Applications of Corpora and Concordancing

The SLA Perspective: Corpus-Based Activities as Consciousness-Raising Activities

Lewis's emphasis on the noticing of lexical and grammatical patterns lends itself to work with authentic texts, both oral and written. When, however, the item in question does not recur frequently in a connected text of manageable size, the use of concordances fills the gap by providing a collection of contextualized occurrences.

It should be noted that Lewis is not alone in proposing the use of concordance-based activities. Barbieri and Eckhardt (2007) argue that the introduction of corpus-based findings in L2 classrooms and instructional materials can and should be informed by current theories of SLA. Accordingly, they point out that corpus-based activities fit well within the parameters of Fotos's (2002) "structure-based interactive tasks" or Ellis's (2003) "focused, consciousness-raising tasks," within the paradigm of planned or proactive form-focused activities (p. 334).[6] Ellis (2003, p. 163) identifies the main characteristics of consciousness-raising (C-R) tasks as follows:

1. There is an attempt to isolate a specific linguistic feature for focused attention.
2. The learners are provided with data that illustrate the targeted feature and may also be provided with an explicit rule describing or explaining the feature.
3. The learners are expected to utilize intellectual effort to understand the targeted feature.
4. Learners may be optionally required to verbalize a rule describing the grammatical structure.

In this conception, the data observed may be oral or written, authentic or contrived, and the operations required of the learner may involve identification of the targeted structure, judgment as to the correctness or appropriateness of the data, or classification of the data into defined categories. "A C-R task constitutes a kind of puzzle which when solved enables learners to discover for themselves how a linguistic feature works" (p. 163).

In his discussion of C-R tasks, Ellis anticipates the objection that activities with a metalinguistic focus do not constitute tasks in the generally accepted sense. Ellis's response is that such activities do, indeed, qualify as tasks because of the requirement that they be interactive in nature—namely, that learners must engage in talk, using their L2 resources, to arrive collaboratively at the task solution. This condition should be kept in mind by those who wish to maximize learner opportunities for negotiation of meaning in the L2. Although some concordance-based activities might be completed by individual learners or might appear to require use of the L1 for group discussion, the opportunity for meaning-focused use of the target language would obviously be lost in such cases.

Ellis cites two rationales for the use of C-R tasks. The first is based on the hypothesis that explicit knowledge facilitates the acquisition of implicit knowledge (2003, p. 163). As stated by Fotos (1994, p. 326), the grammar C-R task "attempts to call learner attention to grammatical features, raising their consciousness of them, and thereby facilitating subsequent learner noticing of the features in communicative input."

The second rationale cited by Ellis comes from the psychological literature on memory and learning and is expressed slightly differently in the work of various researchers. For example, Craik and Lockhart (1972) make the claim that more significant learning results from "greater depth of processing," whereas Bourke (1996) touts the effectiveness of "discovery learning" through problem solving. Similarly, the branch of learning theory known as constructivism has recently attracted increased interest from educational technologists, as developments in informational technology have greatly expanded the quantity and accessibility of data sources available for data-driven learning activities. Cobb (1999, p. 2) contrasts the two types of knowledge believed to be created by the traditional transmission of knowledge versus the construction of knowledge by the learner as "inert, easily forgotten, and untransferable" in the first case and "memorable and transferable to novel contexts" in the second. He reports confirming evidence of this hypothesis from a study that compared a group of students in an English for Academic Purposes class who used concordance data to construct their own definitions of words with another group who learned dictionary definitions.

On the basis of the preceding rationales, it appears that FL educators should give greater prominence in curricula and classroom activities to lexical patterning and should seek better integration of grammatical and lexical aspects of the L2, rather than continuing to rely on the "slot-and-filler" approach with its predominant emphasis on grammatical structures into which single lexical items are to be inserted. Moreover, language program directors and classroom teachers are encouraged to explore ways in which corpora and concordancing

can be employed to enhance instruction of both lexical and grammatical aspects of the given L2.

Pedagogical Uses of Concordancing

Classroom applications of concordancing may take one of three forms: (1) as a reference or research tool for teachers and advanced learners; (2) for teacher-directed student research; or (3) for the preparation of concordance-based learning and/or assessment activities. Because the actual use of ordinary concordancers and analysis of the resulting data require fairly high-level linguistic skills, these activities are most likely to be successful with learners at advanced levels of proficiency or with exceptionally motivated or linguistically sophisticated learners at lower levels. However, such activities can be introduced at lower levels of instruction, provided they are accompanied by the kind of adaptation of the concordancing tools and database exemplified in Cobb (1999). As mentioned earlier, any ordinary Web browser can also be used as a concordancer, although interpretation of these results requires awareness of the different kinds of text types represented and concomitant implications. For students to make effective use of concordancing for research or reference, they must be trained and guided in the use of the selected tools. Chambers (2005) explores the integration of corpus consultation into course design and describes her own experience and her students' experience with such an initiative in an undergraduate language course. Aston (2001a) also offers a number of descriptions of various pedagogical approaches to the use of corpora, of which Gavioli's (2001) contribution, "The Learner as Researcher: Introducing Corpus Concordancing in the Classroom," is particularly pertinent.

This section focuses on the third category mentioned earlier—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 data-driven learning. In particular, it highlights the use of authentic data and the self-discovery of regularities by the learner. Before considering some examples of such activities, it will be useful to review the principal kinds of units and patterns that can be studied with the aid of a corpus and concordancer.[7]

Lexical Features

    1. Lexical items (words)
- Neologisms: Eng. *wimp*, Fr. *courriel* ("e-mail").
- Polysemous words: nominal and verbal uses of Eng. *lead*, different uses of Fr. *tirer* ("to drag, pull," "to draw [conclusions]").
- Lexico-grammatical features, or word grammar: reflexive versus non-reflexive uses of Fr. *sentir* ("to feel"); Fr. *penser* ("to think") *à* + noun/infinitive, *penser* + infinitive, *penser que* + clause, *penser de* + noun.

    2. Collocations and set expressions
- Verb + nominal complement: Eng. *to draw (a line/distinction/laugh),* Fr. *tirer* ("to pull, draw") *(parti/profit/des conclusions/une leçon).*

- Prepositions occurring with particular types of noun phrases: Fr. *en* + year/name of month/expression of duration/feminine place names
- Idiomatic phrases: Fr. *montrer patte blanche* ("to show white paw" = "to give proof of one's identity")

3. Derivational morphology (to which types of word does an affix attach, and what are its meanings?)
   - The prefix *anti-* (Eng. or Fr.)
   - The nominal suffix *-ment* (Eng. or Fr.)

4. Easily confused lexical items (distinct items that present some similarities)
   - Homonyms: Eng. *since* (conjunction versus preposition), Fr. *leur* ("to them," "their") (pronoun versus determiner)
   - Synonyms (or so-called "translation synonyms"). Examples: Eng. *over* and *above;* Fr. *amener, apporter, ramener, rapporter, faire venir*, all related to Eng. "to bring"; *briser, casser, rompre*, related to Eng. "to break."
   - Paronyms (items of similar form but different meaning). Examples: Eng. *interested* and *interesting;* Fr. *accroissement* ("increase") and *croissance* ("growth"); Fr. *avoir la chance de* + infinitive ("to be lucky enough to . . .") versus *avoir des chances de* + infinitive ("to have a chance to . . .").

Grammatical Features

With respect to grammatical items, there is the fundamental problem of determining the optimal search terms to retrieve all (or as many as possible) of the targeted forms, and only those forms. Unless one is working with a tagged corpus—that is, a corpus in which the syntactic category of each word has been indicated in the text so that it can be read by the concordancer—grammatical searches will be limited to forms or structures that are consistently represented by a small set of lexical items. In any case, when concordancing for pedagogical rather than research purposes, one is usually obliged to work with a limited subset of the total occurrences of a particular structure; such limitations can be justified, provided the subset is generally representative with respect to the pedagogically pertinent features.

Some examples of grammatical structures that lend themselves to concordancing are, for English, the modal *should* or the complementation pattern *to . . . ing.* In French, if the concordancer allows searches for discontinuous features, the restrictive adverbial expression *ne . . . que* ("only") can be examined for its uses and especially its variable position and scope within the clause. Grammatical forms having many different realizations can sometimes be located by searching for associated items. For example, a good number of French subjunctive forms could be found by searching for *il faut que* ("it is necessary that"), *à moins que* ("unless"), or *bien que* ("although"). The converse approach, if one were investigating the various contexts that trigger subjunctive use, would be to search for very frequent subjunctive forms such as the present-tense first-person singular and third-person forms of *être* ("to be").

As with lexical items, concordance-based activities can highlight grammatical structures that are distinct but similar in some respect, whether the similarity

be formal, functional, or semantic. For example, comparisons can be made of the simple versus periphrastic future in English or French, or of English present perfect constructions with *for* and *since.* For French, concordance excerpts can demonstrate the semantic differences between *passé composé* (compound past) and imperfect tenses of verbs such as *vouloir* ("to want"), *connaître* ("to know") and *avoir* ("to have"), where the aspectual distinction between the two tenses leads to differing translations in English.

Finally, where different corpora of appropriate sources are available, concordances can facilitate comparison of the same grammatical feature in different text types. For example, in French, concordance data are particularly effective in revealing differences between informal speech and formal written discourse with respect to negation or variant interrogative forms.

Formats of Concordance-Based Activities

Most often, learners are asked to study concordance data prepared by the teacher and to draw generalizations about the form or the meaning, or both, of the targeted item or items. A simple activity would employ a short concordance of one lexical item or lexical phrase and ask students to infer the meaning of the expression. Such an exercise can be done rather quickly (with the use of an overhead transparency or PowerPoint presentation) and is an effective means of developing learners' ability to discern meaning from context. This approach can also 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 conjectures. Once the students have formulated a hypothesis on the basis of several contexts, the hypothesis can be verified by reference to a dictionary definition.

More complex tasks will require learners to draw generalizations about structure or meaning where the concordance data contain multiple patterns, either structural or semantic, or, more often, both. In such cases, the task can be simplified by presorting the data. Alternatively, learners can be asked to sort the concordance lines into categories that are given to them or that they will define. These tasks work well when carried out in 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 are 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 only one pattern. 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, thereby 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 key word. Other exercise types have as their purpose the reinforcement or assessment of

the learner's acquisition of a distinction or set of patterns that have previously been the focus of a concordance-based discovery activity. These activities are typically of the gap-filling type, but they employ only authentic concordance data. More creative sentence completion tasks, where the part of the sentence that is supplied closely resembles part of a target pattern, may also be used for this purpose.

Sample Concordance-Based Activities

All the sample activities provided here are relatively abbreviated due to space limitations; the actual length of concordances will vary according to the targeted item, the nature of the task, and available time. Example 1 presents an activity requiring students to distinguish between homonyms, in this case the two senses of *since,* causal and temporal (as suggested in Aston, 2001b, p. 20). A follow-up discussion can focus on the kinds of constituents that tend to come after the word in each of its senses. Although the data reveal clear tendencies for each sense, they also show that the type of structure may be the same even when the meaning is different (see line 6). The concordance has also been constructed to include an example of temporal *since* plus a gerund, and the common expression *ever since* without a complement.

> Example 1
> Activity on homonyms: Eng. *since*
> The word *since* has two principal meanings: (1) a temporal meaning, referring to time, and (2) a causal meaning, introducing an explanation. Decide which of the following lines illustrate the temporal sense, and which the causal sense of *since*.
>
> 1. in terms of the crisis of tradition itself. <u>Since</u> a civilizational crisis involves also a crisis
>
> 2. that, and treatment had little variation. <u>Since</u> a fall or blow might have caused it, a cold
>
> 3. 1140 7 for ornaments and decorations. <u>Since</u> Biblical times, rosaries have been made from
>
> 4. the other hand, a little later on he says: "<u>Since</u> 1692 a great but superficial change has wipe
>
> 5. and has been renewed at five-year intervals <u>since</u> 1925. A little farther along the road you
>
> 6. really quite realistic. Many years have passed <u>since</u> a Metropolitan audience heard anything
>
> 7. degree of national political significance. <u>Since</u> a national interpretation cannot be avoided
>
> 8. is not a Saturday, Sunday, or legal holiday. <u>Since</u> April 15, 1962, is on Sunday your return for
>
> 9. I knew only a Protestant-dominated country. <u>Since</u> arriving here, however, I have formed a far
>
> 10. shed in 1927, has been selling steadily ever <u>since</u>. As Sandburg said at the time: "It is as

Whereas Example 1 highlights the relation between a word's meaning and its grammar, Example 2 presents an abbreviated concordance for an activity more strictly focused on the lexical aspects of a word's use. Here learners can be asked to identify the various meanings of Fr. *projeter* ("to project"), 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." Discussion can bring attention to the nature of the relationships between the various meanings—relationships that are, of course, generalizable to many other lexical items.

Example 2
Sample concordance for study of lexical item: Fr. *projeter* ("to project") (*Source: Lexiqum*)

1. le 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. On

2. 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. Presque les deux tiers des répondants (63 pour cent) sont abonnés au

3. Une camionnette circulant en direction ouest, sur cette rue, aurait alors happé la voiture pour la projeter contre l'autobus de la STCUM qui circulait dans la direction opposée. Le chauffeur de l'autobus a

4. libre-échange avec les États-Unis, qui modifie le tracé est-ouest des échanges économiques pour les projeter dans une direction nord-sud. Nombreux sont d'ailleurs ceux qui voient dans cet accord commercial un

5. de bonnes adresses? G.R. Saint-Léonard Avec un tel budget - $4000 par couple - , vous pouvez projeter un séjour qui sera fort agréable. Au mois d'avril, vous trouverez probablement des billets d'avion

The activity of Example 3 focuses on the two most common (and most commonly confused) relative pronouns in French and intends to lead learners to the generalization that *qui* ("who, which, that") is used when the pronoun corresponds to the subject of the relative clause, whereas *que* ("whom, which, that") is used for direct objects. Much "doctoring" has gone into the creation of these concordance samples to eliminate occurrences of *qui* and *que* with other grammatical functions (for example, interrogative pronouns). Concordance lines are also selected so as to include antecedents of both human and inanimate references in both cases, to rule out the common misconception that the distinction is based on antecedent animacy as in English. This activity could be used with the first explicit instruction on relative pronouns, usually late in the introductory course, or in the intermediate or advanced course for recycling and review. Here the task is somewhat simplified by the use of a corpus of spontaneous conversation (with some slight editorial modifications), which generally makes the concordance lines more easily interpretable by beginning learners.[8] While instructions to the learner here and in the following examples are given in English, they could be given in the L2 instead, depending on the proficiency of the

learners and the nature of the task and the intended language of its performance. A similar data set, but where the two pronouns are mixed or deleted, can serve as a subsequent application exercise.

Example 3
Grammar-focused activity: Fr. relative pronouns *qui* and *que*
Compare the examples of relative pronoun *que* in 1–5 with those of *qui* in 6–10. Can you formulate a rule that explains the difference between these two?

1. la cuisine est la seule partie de la maison que j'aimerais bien éventuellement refaire
2. trois mille, euh moi, la classe que j'avais, euh bon ben, il y avait des gens, aussi je vais
3. alors euh moi les quelques points que j'ai eus, c'était grâce au sourire, j'ai beaucoup souri
4. alors je vais, euh le le restaurant que j'ouvrirai donc, le jour où je me lancerai là-dedans
5. quand c'est une classe, il y a des gens que tu apprécies et puis il y en a d'autres que tu
6. dans le pays même tu achètes un plat qui a cette forme un peu d'entonnoir
7. c'est un livre qui a été tourné en film. Le livre a gagné le prix Nobel.
8. oh jamais, mais quelqu'un qui a juste quelques kilos de trop n'ose pas mettre un short
9. parce que j'avais invité des gens qui avaient bu et fumé. Vraiment personne était ivre
10. tu sais, c'est comme la numismatique, les gens qui collectionnent. Bon ben là, je sais pas

An example of a problematic pair of translation synonyms in French are the prepositions meaning "in": *en* and *dans*. This pair lends itself well to a comparative activity at the intermediate level or even the introductory level of instruction. Although "doctored" concordances prepared and printed by the teacher can be used for this purpose, it would be more effective to have students work with one of the user-friendly online concordancers. 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; this fact is conveyed very strikingly by the comparison of two long concordances for the pair. 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 which kinds of words each one commonly occurs. At lower levels of instruction, this activity will prompt learners to notice the use of *en* with months, years, and certain geographical names. In addition, learners can be asked, after some introduction to and familiarization with the concept of set phrases (fixed expressions), to determine if any may be found among the collocations. This question should lead to the observation that *en* occurs in a great number of fixed expressions, whereas *dans* occurs in relatively few.

Finally, if working with a raw concordance, learners should be led to observe that *en* 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.

The concordance given in Example 4 targets the distinction between reflexive and nonreflexive use of the French verb *sentir* ("to feel"), thereby illustrating the relationship between a lexical item's meaning and its grammar. Students should arrive at the generalization that the nonreflexive (transitive) use occurs with a direct object complement, whereas the reflexive (intransitive) use occurs with adverbs (or prepositional phrases that function like adverbs), or with adjectives or adjective phrases. Also evident is the ambiguity of *sentir,* which can mean "to smell" as well as "to feel."

> Example 4
> Activity on word grammar: Fr. *sentir* versus *se sentir* ("to feel") (*Source: Lexiqum*)
> Identify the reflexive and nonreflexive uses of the verb *sentir*. Which kinds of complements occur with these two uses? (A complement is the word or phrase that accompanies a verb, usually following the verb, and completes the meaning of the verb.)
>
> 1. Ensuite, après le déjeuner, nous nous couvrons de crème aux agrumes pour **sentir** le pamplemousse: cette odeur nous donne toujours un air bronzé.
> 2. Je ne pensais pas qu'elle pouvait **sentir** ma colère.
> 3. Et voilà, nous retournons dans l'igloo où nous pouvons voir nos ombres sur la glace, où nous pouvons **sentir** notre souffle, . . .
> 4. J'ai fini par apprendre qu'elle venait ici pour ne pas **se sentir** trop seule, qu'elle aimait les livres et le silence, . . .
> 5. Des exercices, de la marche, des séances de méditation m'aideraient à **me sentir** plus à l'aise.
> 6. J'aime **sentir,** sous mes pieds, la butte de pierre et de terre à la gloire du dieu, pleine et dure: pas de creux, . . .

Another example of word grammar that regularly causes difficulty for learners of French is the expression *il s'agit de* + noun phrase or infinitive. This expression, consisting of the impersonal or semantically empty pronoun *il* ("it") and the reflexive verb *s'agir (agir* "to act"), is usually glossed as "it is a question of"; however, the literal translation of the latter phrase, *il est question de*, is not pragmatically equivalent to *il s'agit de*. A brief look at several examples of *il s'agit de* in context gives a much better understanding of the discourse function of this expression than the English gloss can afford. In Example 5, 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 nonreferential 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 new information usually entails some evaluative judgment of the event or situation.

Example 5

Study of discourse function: Fr. *il s'agit de*

Study the following examples of the use of *s'agir de*. How is this expression used? What function does *il s'agit de* serve in each text?

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 Philadephie (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.

Example 6 contains a matching exercise where the learner must match left and right contexts of the keyword expression. The principle underlying this type of exercise is similar to that observed with structured input activities: Acquisition of the targeted expression is sought by engaging the learner in meaningful processing of a series of utterances all employing that expression. In addition, like all concordance-based activities, the exercise provides practice in making use of contextual clues to construct meaning from a written text. A follow-up discussion could also focus on the distinction between the two meanings and the concomitant difference in verbal mood (indicative versus subjunctive).

Example 6

Sentence completion matching exercise: Fr. *de sorte que* ("in order that", "so that")

De sorte que: Find the ending of each sentence.

1. Ces modèles nous ont donné une très bonne idée de l'endroit où allait frapper Hugo,

2. Des grèves régionales prendront le relais: six par semaine,

3. Ce que les médecins appellent l'obésité abdominale est un facteur de risque très important

4. Les assaillantes ont pris la fuite avant l'arrivée de la police,

5. . . . 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 l'enfant puisse la peler facilement.

    c. . . . <u>de sorte que</u> les gens qui ont un tablier c'est-à-dire une bedaine
       pendante sont plus sujets aux crises cardiaques
    d. . . . <u>de sorte que</u> tout le Québec sera touché à tour de rôle.
    e. . . . <u>de sorte que</u> les autorités ont pu intervenir à temps et prendre les
       mesures d'évacuation qui s'imposaient,

Finally, for an example of a grammar-focused activity that involves extended observation and analysis, the reader is referred to Tribble and Jones (1997, pp. 41–42), who include as one of their sample activities a worksheet on reported speech in English. Learners work with concordance data for both *said* and *told*, first observing the difference in complementation (*said <u>to</u> someone* versus *told someone*). Then the learners are asked to record in a table the main verb following *said* or *told*. Next, they "work out what each speaker actually said" in each case and write this information in the same table. Finally, they compare the two columns and formulate generalizations about how the verb tense has changed from the direct quotation to the reported speech version.

Conclusion

The use of corpora and concordancers is a methodology that deserves wider consideration on the part of FL educators. Data-driven learning can enable language educators to meet several objectives espoused by the profession:

- It introduces authentic language samples as the object of study.
- It takes the burden of responsibility off the teacher to be the language "expert" and places greater responsibility on the learner to discover how the language works.
- It can help remedy the imbalance created by decades of overemphasis on the learning of grammatical structure while severely limiting the targeted lexical repertoire.
- It provides a vehicle for the kind of consciousness-raising activities believed to be useful in communicative approaches.

The use of corpora and concordancing also merits attention because the field is showing increased interest in tools and methods that will enable more learners to reach advanced levels of proficiency. In addition, many of the benefits of data-driven learning can be retained at earlier stages of learning through teacher-devised activities based on modified concordance output; alternatively, the analysis of concordance data can be facilitated by the use of corpora consisting of linguistically less complex texts, such as texts intended for younger audiences or for L2 learners.

Concordance-based activities can facilitate acquisition of both lexical and grammatical forms. With respect to grammatical items, observation of data and hypothesis formulation can help learners focus on the role of grammatical forms in producing differences of meaning, thereby conveying the fundamental principle that communicating effectively is not primarily a matter of following rules but rather an issue of making appropriate choices corresponding to alternative meanings. Shelly (1993) suggests a similar 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.

This chapter closes with a few words from Lewis's (1993) description of the teacher's tasks in facilitating data-driven learning:

> 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, 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 (p. 149).

Thus, rightly employed, data-driven learning is student centered in a fundamental way that goes beyond simply having students communicate about themselves and their lives. Given all the competing demands on the FL curriculum, program directors will need to consider how and to what extent corpora/concordance-based activities can be effectively incorporated into the courses that their TAs teach. In any case, if we are serious about creating effective learning conditions for students, we should not be satisfied with relying on traditional methods for explaining and analyzing grammar. Ideally, more second language educators will accept the challenge to explore the potential benefits of data-driven learning.

Notes

1. See "Grammar Safari" at http://www.iei.uiuc.edu/student_grammarsafari.html for a learner-friendly presentation of simple techniques for using the Internet as a concordancer.

2. Widely used programs include Michael Barlow's MonoConc Pro, available from Athelstan (http://athel.com/); Mike Scott's WordSmith Tools, published by Oxford University Press (http://www.lexically.net/wordsmith/); R.J.C. Watt's Concordance (http://www.concordancesoftware.co.uk/); and TextSTAT, a freeware program from Matthias Hüning (http://www.niederlandistik.fu-berlin.de/textstat/software-en.html).

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, pp. 24–25.

4. I recommend the *Time* magazine corpus created by Mark Davies at Brigham Young University (http://corpus.byu.edu/time/).

5. My description of these activities is adapted from the summary of Moudraia (2001).

6. Katz and Blyth (2007, pp. 35–38) also recommend the use of consciousness-raising activities to make grammar learning a more interactive process. Their book includes numerous examples of consciousness-raising activities for the teaching of the most problematic grammatical features of French.

7. The following inventory of features is loosely based on Aston's discussion of pedagogical concordancing (2001b, pp. 16–19).

8. The corpus in question, the Minnesota Corpus, is available on request from the author.

References

Anderson, B. (2007). Pedagogical rules and their relationship to frequency in the input: Observational and empirical data from L2 French [Electronic version]. *Applied Linguistics, 28*(2), 286–308.

Aston, G. (2001a). *Learning with corpora*. Houston: Athelstan.

Aston, G. (2001b). Learning with corpora: An overview. In G. Aston (Ed.), *Learning with corpora* (pp. 7–45). Houston: Athelstan.

Barbieri, F., & Eckhardt, S. (2007). Applying corpus-based findings to form-focused instruction: The case of reported speech [Electronic version]. *Language Teaching Research, 11*(3), 319–346.

Belz, Julie A. (2005). Corpus-driven characterizations of pronominal *da*-compound use by learners and native speakers of German. *Unterrichtspraxis/Teaching German, 38*(1), 44–60.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge, UK: Cambridge University Press.

Biber, D., Davies, M., Jones, J., & Tracy-Ventura, N. (2006). Spoken and written register variation in Spanish: A multi-dimensional analysis. *Corpora: Corpus-Based Language Learning, Language Processing and Linguistics, 1*(1), 1–37.

Biber, D., Johansson, S., Leech, G., & S. Conrad. (1999). *Longman grammar of spoken and written English.* London: Longman.

Bourke, J. (1996). In praise of problem solving. *RELC Journal, 27,* 12–29.

Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology, 9,* 111–125.

Cobb, T. (1999). Applying constructivism: A test for the learner as scientist. *Educational Technology Research & Development, 47,* 15–31. Citations are from the author's pre-publication version, retrieved September 1, 2007, from http://www.lextutor.ca/cv/Constructivism.html.

Cobb, T. (2007). *Corpus concordance French: The compleat lexical tutor* [Online corpus and concordancer]. Montreal: T. Cobb. http://www.lextutor.ca/concordancers/concord_f.html.

Cobuild. (1990). *Collins–Cobuild English grammar*. London: Collins.

Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? [Electronic version]. *TESOL Quarterly, 34,* 548–560.

Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

Davies, M. *Time corpus of American English, 1923–2006* [Online corpus and concordancer]. (2008). http://corpus.byu.edu/ time/.

Di Vito, N. O. (1997). *Patterns across spoken and written French: Empirical research on the interaction among forms, functions, and genres.* Boston: Houghton Mifflin.

Dodd, B. (2006). *Working with German corpora*. New York: Continuum International Publishing Group.

Ellis, R. (2003). *Task-based language learning and teaching.* New York: Oxford University Press.

Fonseca-Greber, B., & Waugh, L. (2003). On the radical difference between the subject personal pronouns in written and spoken European French. In P. Leistyna & C. Meyer (Eds.), *Corpus analysis: Language structure and language use* (pp. 225–240). Amsterdam: Rodopi.

Fotos, S. (1994). Integrating grammar instruction and commmunicative language use through grammar consciousness-raising tasks [Electronic version]. *TESOL Quarterly, 28*(2), 323–351.

Fotos, S. (2002). Structure-based interactive tasks for the EFL grammar learner. In E. Hinkel & S. Fotos (Eds.), *New perspectives on grammar teaching in second language classrooms* (pp. 135–154). Mahwah, NJ: Lawrence Erlbaum.

Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston (Ed.), *Learning with corpora* (pp. 108–137). Houston: Athelstan.

Grammar safari. (n.d.). Urbana, IL: Intensive English Institute, University of Illinois at Urbana–Champaign. http://www.iei.uiuc.edu/student_grammarsafari.html.

Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English.* Philadelphia: Benjamins.

Katz, S., & Blyth, C. (2007). *Teaching French grammar in context: Theory and practice.* New Haven, CT: Yale University Press.

Labeau, E. (2006). French television talk: What tenses for past time? *International Journal of Corpus Linguistics, 11*(1), 1–28.

Lewis, M. (1993). *The lexical approach: The state of ELT and the way forward*. Hove, UK: Language Teaching Publications.

Lewis, M. (1997). *Implementing the lexical approach: Putting theory into practice*. Hove, UK: Language Teaching Publications.

Lexiqum [Online corpus and concordancer]. (n.d.). Montreal: RALI, Université de Montréal. http://retour. iro.umontreal.ca/cgi-bin/lexiqum.

Lutzeier, P. R. (1999). Das "Wort"-Korpus: Fragen und Antworten zu seiner Rolle am Beispiel der Präposition *um*. *Deutsche Sprache: Zeitschrift für Theorie, Praxis, Dokumentation, 27*(2), 118–133.

Marnette, S. (2003). Indirect discourse in a diachronic and pangeneric corpus of French: Forms and strategies. *Forum for Modern Language Studies, 39*(4), 414–426.

Möllering, M. (2004). *The acquisition of German modal particles: A corpus-based approach*. Bern, Switzerland: Peter Lang.

Moudraia, O. (2001). *Lexical approach to second language learning* (Report No. EDO-FL-01-02) [Electronic version]. Washington, DC: ERIC Clearinghouse on Language and Linguistics, Center for Applied Linguistics. Retrieved September 8, 2007, from http://www.cal.org/resources/digest/ 0102lexical.html.

Nattinger, J., & DeCarrico, J. (1989). Lexical phrases, speech acts and teaching conversation. In P. Nation and R. Carter (Eds.), *AILA review 6: Vocabulary acquisition* (pp. 118–139). Amsterdam: AILA.

Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford, UK: Oxford University Press.

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, UK: Cambridge University Press.

Polzin-Haumann, C. (2003). Grammaticographie contre usage? A propos de la norme du pronom relatif *dont* en français contemporain. In D. Osthus, C. Polzin-Haumann, & C. Schmitt (Eds.), *La norme linguistique: Théorie–pratique–médias–enseignement* (pp. 99–114). Bonn, Germany: Romanistischer Verlag.

Porto, M. (1998). Lexical phrases and language teaching [Electronic version]. *English Teaching Forum, 36*(3), pp. 22–26.

Ramón García, N. (2002). Semántica y sintaxis: El sintagma nominal en un estudio contrastivo inglés–español. *Miscelánea: A Journal of English and American Studies, 25,* 143–159.

Sancho Sánchez, M. (1999). Frecuencia y gradación de algunas estructuras subordinadas en español. *Cuadernos Cervantes de la Lengua Española*, 5(25), 42–45.

Shelly, S. (1993). Rule reformulation at the advanced level. *French Review, 66,* 760–768.

Sinclair, J. (1991). Corpus, Concordance, Collocation. Oxford, UK: Oxford University Press.

Sinclair, J. (2000). Lexical grammar [Electronic version]. *Naujoji Metodologija*, *24*, 191–202. Retrieved September 23, 2007, from http://donelaitis.vdu.lt/publikacijos/sinclair.pdf.

Stenström, A.-B. (2006). The Spanish pragmatic marker *pues* and its English equivalents. In A. Renouf & A. Kehoe (Eds.), *The changing face of corpus linguistics* (pp. 263–281). Amsterdam: Rodopi.

Tribble, C., & Jones, G. (1997). *Concordances in the classroom: A resource guide for teachers.* Houston: Athelstan.

Woolard, G. (2000). Collocation: Encouraging learner independence. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 28–46). Hove, UK: Language Teaching Publications.

Zinggeler, M. (2006). "GRIMMATIK": German grammar through the magic of the Brothers Grimm fairy tales and the Online Grimm Corpus. In E. Fitzpatrick (Ed.), *Language and Computer: Vol. 60. Corpus Linguistics Beyond the Word: Corpus Research from Phrase to Discourse* (pp. 167–184). Amsterdam: Rodopi.

Zinggeler, M. (2007). *GRIMMATIK: German grammar through the magic of the Brothers Grimm fairy tales. LINCOM studies in German linguistics* (vol. 26). Munich: LINCOM.

Appendix: Online Corpus and Concordancer Sites

French

The Compleat Lexical Tutor. UQAM (Université du Québec à Montréal) Web Concordancer authored by T. Cobb, v. 6, November 2007. http://www.lextutor.ca/concordancers/concord_f.html

Easy-to-use and fairly versatile concordancer, with a choice of several different corpora, including all of *Le Monde* from 1998. Also includes a modest corpus of spoken French paired with a written corpus of equal size.

Lexiqum. Montreal: RALI, Université de Montréal. http://retour.iro.umontreal.ca/cgi-bin/lexiqum

User-friendly, versatile site with extensive corpora of literary, journalistic, and administrative Canadian texts.

ELICOP. Etude Linguistique de la Communication Parlée, Département de Linguistique, Université Catholique de Louvain (Belgique). http:// bach.arts.kuleuven.ac.be/elicop/

Extensive corpora of spontaneous spoken French, including 80 hours of the Orléans corpus (1968–1971). Given the "bare" nature of the transcriptions (no punctuation), their deciphering is a challenge for students. There are several different versions of the concordancer with different entry points; one allows syntactic searches of this tagged corpus.

ARTFL. Project for American and French Research on the Treasury of the French Language, University of Chicago. http://humanities.uchicago.edu/ARTFL/ARTFL.html

A huge searchable database of texts from fifteenth- to twentieth-century French literature, philosophy, arts, and sciences. Can do simple or sophisticated searches, but requires a little learning.

German

COSMAS. Corpus Search, Management and Analysis System, Mannheim Institut für deutsche Sprache. http://corpora.ids-mannheim.de/ccdb/

Publicly available portion of the world's largest collection of German online corpora. Wide variety of sources, including some spoken language, totaling more than 1.1 billion words. Searches limited to 60-minute sessions.

German National Corpus. DWDS (Digitales Wörterbuch der deutschen Sprache des 20. Jh.). http://www.dwds.de/cgi-bin/rest/loginstart

The core corpus, comprising a range of written and spoken text-types, is available for online search at this site after free-of-charge registration.

Spanish and Portuguese

Corpus del Español and Corpus do Português. Mark Davies, Brigham Young University. http://corpus.byu.edu/

Everything you could need, in one place. Two large, comprehensive corpora of literary texts from the medieval to the modern period. The additional **Corpus del Español: Registers** is a smaller but still large corpus equally divided between fiction, nonfiction, and spoken texts. Tagged corpora with versatile and user-friendly concordancers.

Multilingual Corpora

Leeds Collection of Internet Corpora. Centre for Translation Studies, University of Leeds. http://corpus.leeds.ac.uk/internet.html

Tagged corpora of Internet texts for 11 languages, including Chinese, Japanese, Italian, Polish, and Russian. Good for informal language.

C-ORAL-ROM Encrypted Corpus. CD distributed with *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages,* E. Cresti and M. Moneglia (Eds.), (2005). Amsterdam: Benjamins.

Four comparable corpora of spoken language only for Italian, French, Spanish, and Portuguese. Each 300,000-word corpus is presented in multimedia format, allowing simultaneous access to aligned acoustic and textual information, as well as keyword searches of texts.