DESIGNING A CORPUS-INFORMED EFL TEXTBOOK

I. P. Dilay,
Ph.D., Associate Professor,
Ivan Franko National University of Lviv,
e-mail: irynadilay@gmail.com
M. P. Dilai,
Ph.D., Associate Professor,
National University "Lviv Polytechnic",
e-mail: marianka_d@yahoo.com

The article concerns itself with enhancing the quality of materials, namely textbooks, which are used for the teaching of foreign languages and English in particular. Teaching English as a foreign language requires authentic language material to be reflected in modern textbooks. Electronic English language corpora can serve as reliable sources of naturally occurring language samples that can be efficiently implemented in textbooks and other language materials.

Corpus-informed teaching and learning as part of computer-assisted language learning can be applied both directly and indirectly. Corpora as repositories of naturally occurring language contain samples of written and spoken communication, reflecting all genres and registers. Exposure to corpus-based materials, though challenging at the beginning, can significantly boost learners' motivation and self-reflection.

Designing a corpus-informed EFL textbook for Ukrainian L1 learners crucially depends on the existence of Ukrainian English learner corpus. Via providing the interface between L1 and L2 learner corpora help identify the errors typical of particular L1 learners and thus suggest appropriate material selection for learning L1. Material selection is normally based on frequency and salience, as well as specific pedagogical tasks. The stages of designing a textbook based on corpus data and some plausible exercises and activities have been suggested in the article.

Corpus-informed language teaching materials can serve as means of interactive learning and be applied at all levels of English language teaching. Thus, there is a need to publicize corpora as globally acclaimed language learning resources to home educators, materials developers, teachers and learners.

Key words: teaching English as a foreign language **(**TEFL); corpus; a corpus-informed textbook; a learner corpus; computer-assisted language learning (CALL).

Problem statement. There is a vast range of textbooks for teaching English as a foreign language, published both in Ukraine and abroad. Textbooks are

generally based on a grammatical syllabus relying either on long-standing tradition of English language teaching or the accepted knowledge found in current handbooks. On studying the existing English language school textbooks it has become clear that the English which is taught in them is at odds with the language used by native speakers.

Modern materials development requires to:

1) clarify the terms and concepts commonly used in discussing materials development;

2) carry out systematic evaluations of materials currently in use in order to find out to what degree and why they facilitate the learning of a language;

3) consider the potential of current research into foreign language teaching;

4) consider the potential applications of what both teachers and learners believe is valuable in the teaching and learning of a foreign language;

5) pool resources and bring together researchers, writers, teachers, learners and publishers in joint endeavors to develop quality materials;

6) to rely on natural language material and expose the learners to authentic samples, both written and spoken, of its use.

Previous research. According to B. Tomlinson, materials development refers to anything which is done by writers, teachers or learners to provide sources of language input and to exploit those sources in ways which maximize the likelihood of intake: in other words the supplying of information about and/or experience of the language in ways designed to promote language learning [1, p.2]. For materials to be valuable the leaning points should be potentially useful to the learners and the learning procedures should maximize the likelihood of the learners actually learning what they want and need to learn.

A survey conducted by MATSDA (The Material Development Association) for TEFL revealed that most of the authors of well-known materials published abroad follow their intuitions rather than an overt specification of objectives, principles and procedures [1, p. 4]. Obviously these intuitions are informed by experience of what is valuable to learners of a language and in many cases do lead to the development of new materials. As a result authors tend to clone the successful features of their competitors. Otherwise, considerable time and expenditure is needed to develop motivating materials.

It has been continuously claimed by language education researchers that materials should expose the learners to language in authentic use. Such authentic language use can be found in modern electronic corpora. The use of corpora for materials development has been rigorously advocated by B. Tomlinson, T. Johns, L. Flowerdew, A. J. Liddicoat, S. Fligelstone,

A. Wichmann, J. Sinclair, S. Hunston et al. They argue that textbooks need to provide exposure to authentic input through the advice they give, the instructions for their activities and both spoken and written texts they include. Ideally materials at all levels should provide authentic input which is rich and varied [1, p. 13]. The input should vary in style, mode, medium and purpose and should contain the features characteristic of authentic discourse in the target language. It is important that the learners are exposed to planned (e. g., a lecture), semi-planned (e. g. an interview) and unplanned discourse (e. g. a spontaneous conversation). The materials should also stimulate learners' interaction with the input and one another rather than just passive reception of it.

Aim and objectives. Based on the modern corpus-based (corpus-informed) approach to materials development, the aim of this article is to reveal the potential of corpora for designing English language textbooks for Ukrainian learners. It presupposes:

1) defining the very notion of a language corpus, its types, structure;
2) elaborating the notion of a corpus-informed (corpus-based) textbook;
3) providing examples of corpus-based exercises and activities which can be included in the textbook.

Discussion. The rapid development of the ability of computers to handle large amounts of language data made it possible to build large language corpora which allow researchers to analyze how the language is being used at the present time (synchronic corpora) or in the past (diachronic corpora). Corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research [5]. Dictionaries, grammars and textbooks compiled before the advent of corpora tended to rely heavily on books that were already published. Corpora enabled to lay aside intuitions and look at the data. The first corpora, which included the Brown Corpus, completed in 1964 (Kuchera and Francis, 1967), the LOB Corpus, completed in 1978 (Hofland and Johansson, 1982), were rather small. Nowadays, the corpora include billions of words and oftentimes Web itself is considered a corpus. It is seeking for the adequate querying procedures and making corpora balanced and representative of all types of language that are the key issue of modern corpus linguistics. The size of the corpus depends on the needs of its users.

Corpora serve as invaluable sources of word frequency information which can pinpoint potentially indispensable vocabulary items and patterns for the language learners. What is noteworthy is that such information can be easily updated each year and integrated into developing modern authentic corpus-informed textbooks. It is obvious that more teaching and learning effort

125

should go into the most frequent senses of words and instances of use. Spelling variations frequencies can also help developers and teachers to decide what to teach first. For instance, 'judgment' is twice as common as 'judgement' and 'inquire' is more common than 'enquire'. At this point the variant or dialect is irrelevant. If a material developer decides to penetrate into language variation, such information is well reflected in modern language corpora, such as the BNC, COCA, COHA, GloWbE, etc. BYU site created by Mark Davies provides access to publicly available English language corpora [7].

Watching movies, videos, documentaries or going straight abroad to learn it is not always affordable for the learners and teachers. A textbook that can help handle typical everyday situations featuring authentic exchanges is an asset to the classroom. For instance, it is also hardly ever mentioned in existing how to handle typical formal situations where the speakers are supposed to use polite addresses to the teacher, doctor, professor, official or even a bus driver. Such information can be also retrieved from corpora, in particular from their spoken (or video) section and implemented in the textbook in the form of its media support.

It is important to know how to select the items to fit particular teaching and learning tasks. Is it enough to rely on frequency only and provide examples from corpora rather than invent them or borrow from other authors? The massive corpus data can look overwhelming for the majority of researchers who have insufficient experience using it. The level of the learners should be taken into account when choosing texts and samples for activities, exercises, instructions, etc. Besides, it has been argued that some language features, salient as they are, are underrepresented since infrequent in corpora. Such salient items are usually culture specific. They might have low frequency in corpora but appear to be relevant to learn.

Corpus-informed or corpus-based learning presupposes the use of novel electronic language corpora to develop modern language learning materials. Mainly it pertains to the use of L2 corpora, English in this case. Nonetheless, it may include parallel corpora with the translations of samples, unfortunately not available for the Ukrainian language, as well as developmental corpora (see CHILDES) as corpora which reflect the language acquisition by children of different age who are English language native speakers. Unfortunately, the respective corpora of L1 (Ukrainian) have not been designed yet. There is, however, one more prerequisite of a successful EFL textbook compilation. It should meet the needs of L1 (Ukrainian) native speakers; namely, take into account their major difficulties in learning English imposed by the structure of their L1. Achieving the balance between what is frequently used, age relevant, salient for L2, and difficult for L1 is the key to success in designing a modern language textbook.

International reported experience draws inferences about L1 difficulties from surveys of the learners. This practice tends to be time and effort consuming and subjective because it is not clear whether it was the teacher or the textbook that failed to get the message across or a particular learning attitude. It is hard for the learners, especially for the young ones, to air their learning problems or admit them. Still it is a valuable praxis and it is worth to ask the learners about how they want their textbooks to improve. They tend to envisage them as more interactive and less theoretical.

Another way to locate learners' difficulties is to conduct an error analysis. It involves collecting large and representative databases – corpora of learners' output, be it written or spoken. Such corpora are called learner corpora and require following special procedure in order to design them. L2 learner corpora are being designed for a number of L1 language learners. ICLE is a project of International Learner Corpus in Louvain, Belgium, whose purpose is to study the differences between English language learners depending on their L1, compare their output with that of the native speakers, find and account for typical errors and make suggestions on how to handle the learners' problems and develop L1 centered materials, including TEFL textbooks. Unfortunately, the Ukrainian contribution to this project is scarce. There is an urgent need to design such learner corpora in Ukraine with the implementation of the state-of-the-art corpus linguistics methodology. Though there are databases of students' tests but they are not electronic and annotated. There are hardly any spoken classroom language databases featuring the students' uncontrolled performance. Learner corpus is on time. Relying on its data, as well as on L2 corpus data can considerably improve the quality of modern EFL textbooks of any level.

Teachers need better training both in developing language materials based on intuition and experience, and in applying corpora for teaching purposes. Corpora can be applied directly and indirectly in the classroom. The direct exposure to corpora, even if it is mediated by the teacher, is at least at the beginning somewhat overwhelming for the learners. It means that the learners become familiar with corpus interface, making queries, can handle the output by themselves. The indirect exposure to corpora requires expertise on part of instructors and material developers when learners can hardly realize what a corpus is. Though, if a textbook is corpus-informed, learners will notice the difference from the traditional textbook content and, if it is properly designed, will acknowledge it.

Corpora come in particularly handy by providing enough exercises to ensure that the student is confronted by a different set of examples [4, p. 117]. Corpus-informed textbooks can relieve teachers of the need to prepare special materials and leave them with more time to devote to learning encouragement, motivation, and fostering group activities and

learning through interaction. It is especially topical at the time when social constructivism language teaching methods (Kramsch, Kasper and Lantolf) focusing on interaction and team work are being prioritized.

It will throw light on the merits of the corpora as sources of language exercises to consider some examples. Wilson [2, p. 119] suggests classifying the exercises by the complexity of search needed to derive the material for them, rather than by their purpose. The exercises fall into lexical, syntactic and discourse. The exercise can be generated automatically from the corpora without human intervention at any point [2, p. 119]. Concordancers (KWIC) can be particularly helpful when designing exercises focusing on particular words, or their senses, collocations and colligation models. Supplementary corpus tools created for the needs of teachers and learners can be leveraged to automatically retrieve concordances, disambiguate word senses with the help of dictionaries and design exercises. For instance, Compleat Lexical Tutor can generate concordances in any language and build the exercises based on them [8]. Learners can benefit both from indirect computer-assisted learning (CALL) via modern corpus-informed textbooks, as well as from the direct exposure to user-friendly online concordancers.

Learners can be given fairly simple tasks involving mark-up of the context to identify collocations, etc [4, p. 101]:

How many different verbs are shown with this structure?

What are the expressions with the pattern 'at_____ time'?

Identify the contexts in which the structure has a clearly positive meaning, and those in which it has a clearly negative meaning. What do you conclude from the results?

Using clues from the context, what can you discover about the places and institutions named in these citations?

Judging from the citations, is this structure used more often in formal or informal varieties of English?

These tasks are more open, in that there may be more than one possible classification or explanation of a phenomenon in the data. In addition, students define their own tasks as they start noticing features of the data for themselves. Exercises may be developed based on gapping and reordering citations. A concordance-based exercise is called OIMC (One Item, Many Contexts) [Johns, 102]. It is highly recommended that the number of citations presented for each item in OIMC should be at most seven.

Working with data leads to a radical revision of preconceived ideas about what one should be teaching as well as how one might teach it. Redefining what to include in textbooks is critical for raising learners' competence and performance. Corpus-informed textbook can suggest ways of dealing with areas of language which have traditionally been poorly taught or regarded as unteachable (e.g. article usage) and reveal the areas of modern structure

(e.g. collocations) that have been neglected before. The data controls not only which features of the language are taught, but which exponents are presented and which senses are taken as primary based on experience. More fundamentally, the traditional division between independent 'levels' of language (lexis – syntax – discourse) appears to be untenable when focusing on word context [2, p. 102].

Thus the following fundamental steps for developing modern English language textbooks can be proposed:

1) contrastive data analysis of an English language corpus and a Ukrainian learner corpus;

2) construction of a didactic material based on corpora;

3) derivation of pedagogical material from the didactic material.

The pedagogic material is based on those features from the didactic material which are central and recurrent. In addition, the selection is determined by the objectives of courses, the age of the learners, their development as well as sociocultural background. Retrieving didactic material from corpora is the initial step of all follow-up work. Based on corpus findings, D. Mindt has argued, for instance, that the alphabetical list of irregular verbs is not the efficient way to learn them. After being exposed to ten most frequently used irregular verbs (*say, make, go, take, come, see, know, get, give, find)* the learner will master 45.6 per cent of the verb patterns of irregular verbs [6, p. 49].

Some researchers emphasize the necessity of corpus planning since its decisions affect the form of the language that is used and valued within educational system [3, p. 993].

Conclusion. The article was an attempt to substantiate the need for modern corpus-informed EFL textbooks in Ukraine. The traditional approach leads to developing textbooks which very often do not adequately reflect the use of present-day English. Corpora geared to foreign language teaching can do much to bring the teaching of English into better accordance with actual use. The major problems concerned with adopting the corpus-informed approach to materials development are lack of training in corpus methodology and popularization of corpora, which is one of the reasons why a Ukrainian learner corpus has not been created so far. In this respect, integration of corpus-based methodology in modern language education in Ukraine appears to be topical and promising.

References

1. Applied Linguistics and Materials Development (ed. by B. Tomlinson). – London, New Delhi, New York, Sydney : Bloomsbury, 2013. – 335 p.
2. Flowerdew L. Corpora and Language Education / L. Flowerdew. – New York : Palgrave Macmillan, 2012. – 347 p.

3. Liddicoat A. J. Corpus Planning: Syllabus and Materials Development / A. J. Liddicoat // Handbook of Research in Second Language Teaching and Learning (ed. by E. Hinkel). – Mahwah, New Jersey : Lawrence Erlbaum Associates, 2005. – pp. 993–1011.
4. Materials Development in Language Teaching (ed. by B. Tomlinson). – Cambridge : Cambridge University Press, 1998. – 368 p.
5. Sinclair J. (2004), How to Use Corpora in Language Teaching / J. Sinclair. Amsterdam : John Benjamins Publishing, 2004. – 239 p.
6. Teaching and Language Corpora (ed. by A. Wichmann, S. Fligelstone et al.). – London and New York : Longman, 1997. – 343 p.
7. Davies M. (2008) The Corpus of Contemporary American English: 520 million words, 1990-present. Available online at www.corpus.byu.edu/coca
8. The Compleat Lexical Tutor. Available at www.lextutor.ca

Ділай І. П., Ділай М. П.
УКЛАДАННЯ ПІДРУЧНИКА З АНГЛІЙСЬКОЇ МОВИ ЯК ІНОЗЕМНОЇ НА ОСНОВІ КОРПУСУ

У статті розглянуто шляхи покращення дидактичних матеріалів, зокрема підручника з англійської мови як іноземної для усіх навчальних рівнів. Наголошено, що навчання англійській мові вимагає представлення у підручниках аутентичного мовного матеріалу, джерелом якого можуть слугувати мовні корпуси. Залучення корпусів як елементу навчання за допомогою комп'ютера у навчальний процес підвищує учнівську мотивацію та саморефлексію.

Підручник на основі корпусу повинен інтегрувати результати відбору матеріалу за частотністю, важливістю та педагогічною релевантністю. Зокрема, використання даних відповідного учнівського корпусу дасть змогу зосередити увагу на типових труднощах та окреслити шляхи їх подолання.

Ключові слова: навчання англійській мові як іноземній; корпус; учнівський корпус; підручник на основі корпусу; вивчення мови за допомогою комп'ютера.

Дилай И. П., Дилай М. П.
СОСТАВЛЕНИЕ ОСНОВАННОГО НА КОРПУСЕ УЧЕБНИКА ПО АНГЛИЙСКОМУ ЯЗЫКУ КАК ИНОСТРАННОМУ

В статье рассмотрено способы улучшения дидактических материалов, в особенности учебника по английскому языку как иностранному для всех уровней обучения. Внимание акцентировано на том, что обучение английскому языку требует наличия в учебниках аутентического языкового материала, источником которого выступают корпуса. Внедрение корпусов как элемента обучения с помощью компьютера в учебный процесс способствует повышению мотивации в учеников.

Учебник, основанный на корпусе, должен интегрировать результаты отбора материала за частотностью, важностью и педагогической релевантностью. Применение данных соответственных учебных корпусов дает возможность сосредоточить внимание на типических трудностях и определить пути их преодоления.

Ключові слова: обучение английскому языку как иностранному; корпус; учебный корпус; учебник, основанный на корпусе; изучения языка с помощью компьютера.

МЕТОДОЛОГІЧНІ ПОНЯТТЯ РОЗДІЛУ "ЕЛЕКТРОДИНАМІКА" В ПІДРУЧНИКУ ФІЗИКИ ПРОФІЛЬНОГО РІВНЯ

Д. О. Засєкін,
кандидат педагогічних наук,
Інститут педагогіки НАПН України,
старший науковий співробітник відділу
біологічної, хімічної та фізичної освіти,
e-mail: dmytro_z@ukr.net

У статті розглядаються принципи побудови структури та добору змісту розділу "Електродинаміка" з метою розкриття його методологічних понять в підручнику фізики профільного рівня. Необхідність формування методологічних знань учнів є важливою ланкою у системі формування предметної та науково-природничої (як ключової) компетентностей учнів.

Ключові слова: методологічні знання; компетентність; підручник.

Постановка проблеми. Шкільний курс фізики побудовано за двома логічно завершеними концентрами: в основній школі (7—9-й класи) вивчається базовий курс фізики, який закладає основи фізичного знання на феноменологічному рівні; у старшій школі вивчення фізики відбувається залежно від обраного профілю навчання. Конструювання змісту і структури другого концентру (10—11-й класи), зокрема профільного рівня, навколо фундаментальних фізичних понять і теорій та обгрунтування відповідної методики й засобів навчання було і залишається актуальною теоретичною і практичною потребою. На необхідності обгрунтування принципів відбору і

131