

Dana-Anca Cehan

ACCRETION: THE CASE OF MLDS

Rezumat

Lucrarea “Accretion: The Case of MLDS” (“Acumulări: cazul dicționarilor pedagogice monolingve pentru străini”) urmărește evoluția unui dicționar tipic pentru lexicografia englezească, ilustrând modul în care lexicografia pedagogică se constituie în interfața dintre lingvistica teoretică, lingvistica aplicată și practica predării limbii engleze ca limbă străină. Lexicografia în general este un domeniu unde se pot evidenția contribuțiile lingvisticii teoretice puse în slujba practicii dar și locul unde se elaborează teorii și principii necesare descrierii limbii. Și cum distincția dintre teorie și descriere nu a fost încă tranșată definitiv, se poate afirma că primii lexicografi au marcat atât debutul lingvisticii teoretice și al celei aplicate cât și cel al predării sistematice a limbilor străine.

În timp ce interesul teoreticienilor limbii se poate limita astăzi la erorile lingvistice sau sociale din perspectiva deviației de la normă, interesul lingvisticii aplicate și a practicienilor se concentrează pe necesitatea existenței unei autorități care să explice ce este acceptabil în funcție de standarde, norme și convenții și să ofere exemple. Și totuși, recomandările au un suport teoretic, bazându-se pe considerații legate de regularizarea limbii și a factorilor care o determină. De fapt, atât atitudinea prescriptivă cât și cea proscriptivă sunt legate de standarde și de cultivarea lor. Ele se combină în dicționarele pedagogice moderne prin integrarea gramaticii și a lexicografiei, făcând ca teoria, descrierea și folosirea limbii să devină aici convergente. Dicționarele contribuie astfel la standardizarea diverselor aspecte ale limbii. Deși în secolul al XX-lea dicționarele și gramaticile au devenit preponderent descriptive, ele nu au putut ignora importanța normelor și a standardelor și astfel și-au păstrat rolul de autorități, așa cum demonstrează piața în permanentă expansiune a dicționarilor, gramaticilor și ghidurilor ortografice și de punctuație.

Lexicografia pedagogică este o activitate cumulativă, cu o istorie milenară care începe cu tăblițele de lut folosite în pregătirea scribilor babilonieni. Lexicografia pedagogică britanică a început în secolul al XVI-lea cu listele de cuvinte publicate de practicienii care încercau să standardizeze ortografia și să explice cuvintele ‘grele’. Cum expansiunea limbii engleze și

dobândirea de către aceasta a statutului de limbă de comunicare globală au transformat în timp studiul acestei limbi într-o activitate mondială, piața materialelor pedagogice a cunoscut o permanentă dezvoltare, beneficiind atât de evoluțiile teoretice cât și de cele tehnologice. Cei patru mari editori britanici – Longman, Oxford, Cambridge și Collins – publică periodic, începând din 1942, dicționare pedagogice monolingve pentru diferite nivele de studiu. Lor li s-au adăugat mai recent editurile americane, canadiene și australiene.

Lucrarea de față descrie locul special ocupat de dicționarul monolingv pentru străini între dicționarele pedagogice. Ea oferă o descriere a elaborării și acumulării lente a convențiilor lexicografice, începând cu dicționarele primilor lexicografi britanici (Robert Cawdrey, John Bullokar, Nathan Bailey și Samuel Johnson) și pune în evidență legăturile strânse dintre teorie, aplicație și practică.

Capitolul dedicat contribuției membrilor mișării ‘Vocabulary Control Movement’ (Harold Palmer, Michael West și A. S. Hornby) analizează primul dicționar pedagogic monolingv pentru străini, *Idiomatic and Syntactic English Dictionary (ISED)* din 1942 și versiunile sale ulterioare: *The Advanced Learner’s Dictionary*, *The Oxford Advanced Learner’s Dictionary of Current English*, *The Advanced Learner’s Dictionary of Current English* și *An Advanced Learner’s Dictionary of English*. Mișcarea a inspirat principiul legăturii strânse dintre structurile gramaticale și lexic și cel al identificării unui vocabular minim bazat pe frecvența utilizării cuvintelor. *ISED* este prima ilustrare a principiilor de bază ale lexicografiei pedagogice anglo-saxone: selectarea cuvintelor definite în funcție de frecvență și utilitate, includerea pronunției standard prin transcriere fonetică, folosirea unor definiții simple și directe, ilustrarea sensurilor prin exemple contextualizate, prezența unor explicații privind deosebirile dintre sinonime și a unor informații gramaticale incluzând structurile verbale, a unor informații privind colocațiile și registrul. Michael West a influențat evoluția dicționarului pedagogic monolingv pentru străini prin publicarea *General Service List of English Words (GSL)*, documentul care conține principiile de selecție a vocabularului bazată pe frecvență, valoare structurală, universalitate, acoperire semantică, flexibilitate și stil. Pe scurt, în perioada pre-computațională, caracteristicile dicționarului pedagogic monolingv pentru străini au evoluat în urma acumulării unor principii încă actuale care i-au conferit unicitate și au contribuit la formalizarea treptată a unei viziuni asupra lexicului și a gramaticii.

A doua perioadă din istoria dicționarilor pedagogice monolingve pentru străini începe în 1987 și coincide cu debutul conlucrării dintre lexicografie și lingvistica computațională și publicarea dicționarilor *Oxford Advanced Learner’s Dictionary of Contemporary English*, *Longman Dictionary of Contemporary English* și, mai cu seamă, *COBUILD*. Lingvistica

compuțională a oferit o nouă perspectivă asupra limbii, o nouă modalitate de descriere a acesteia și teorii noi de învățare și însușire a ei, care folosesc date și concepte noi. O dată cu schimbarea tipului de date analizate de la cele intuitive la cele naturale, studiul corpusurilor a reorientat interesul lingvisticii și al predării limbilor dinspre cunoașterea limbii către folosirea acesteia. Deși analiza empirică a datelor naturale nu este o inovație a lingvisticii compuționale, bazele de date tot mai extinse ale anilor '80 au început să pună în evidență discrepanțe între rezultatele analizelor de date provenind din folosirea reală a limbii și cele ale studiilor bazate pe introspecție. Corpusurile masive și noile tehnologii de prelucrare a datelor au înlesnit descrieri mai precise și mai amănunțite și au pus în evidență nu numai regularități și tendințe care nu fuseseră intuite dar și structuri care nu fuseseră detectate până în acel moment. În plus, diferențele dintre limba vorbită și cea folosită în scris au putut fi documentate mai convingător (e.g., Biber *et al.* 1999; Carter și McCarthy 2006).

Anul 1987 a fost punctul de cotitură al lexicografiei pedagogice britanice prin publicarea de către John Sinclair a dicționarului *COBUILD (Collins COBUILD English Language Dictionary)*, un dicționar inovativ, care se bazează exclusiv pe analiza unui corpus masiv, conține exclusiv exemple autentice și folosește definiții sub formă de fraze sau propoziții complete. Cuvintele definite sunt încadrate în frazeologia lor tipică, ilustrând astfel dependența dintre uz și sens, iar menționarea cologațiilor caracteristice ilustrează principiul conform căruia orice diferență de sens se asociază cu o diferență formală. De fapt, *COBUILD* a fost primul dicționar care a demonstrat că structurile co-selectate afectează înțelesul cuvântului sau că sensul este o proprietate a sintagmelor în aceeași măsură în care este o proprietate a cuvântului. Sinclair urmează principiile formulate de 'Vocabulary Control Movement', combinate cu avantajele oferite de dezvoltarea tehnologiei informației, iar *COBUILD* pune accent pe explicarea folosirii cuvintelor și nu pe inventarierea sensurilor și ilustrează modul în care fiecare sens se asociază cu o structură specifică de folosire. În plus, începând cu *COBUILD* frecvența de folosire a cuvintelor este marcată prin diferite simboluri, alături de informația semantică, pragmatică și cea gramaticală. O altă inovație lexicografică lansată de *COBUILD* sunt definițiile constând din propoziții sau fraze complete, formulate în termeni familiari și asemănătoare exemplilor. În pofida criticilor pe care le-a întâmpinat, *COBUILD* rămâne unul dintre cele mai influente dicționare pedagogice ale tuturor timpurilor.

După *COBUILD*, lexicografii au renunțat treptat la ideea că un cuvânt izolat poate avea mai multe sensuri și la cea că definițiile de dicționar sunt precise și se exclud reciproc. Patrick Hanks a arătat că un cuvânt izolat are doar potențiale semantice, iar sensurile lui provin din sintagmele în care este folosit. Împreună cu Pustejovsky, Hanks a demonstrat că sensurile unui

cuvânt nu se exclud și că nu există criterii sigure pentru a distinge un sens de altul și că, prin urmare, sensul este vag, imprecis și probabilistic. Limba însăși este un sistem probabilistic și preferențial în care contextele se corelează cu structuri relevante statistic și cologații. Aceste preferințe și probabilități pot fi prezentate doar într-un nou tip de dicționar, bazat pe construcții. Cu aceste noi teorii, se pun bazele unor noi principii de elaborare a dicționarilor, care conduc spre o nou tip de dicționar, cel bazat pe asocieri între sens și construcție și, probabil, o nouă fază în evoluția lexicografiei.

Intre timp, a început să fie construit un nou tip de corpusuri, corpusurile pedagogice ('computer learner corpora'), care conțin texte create în contexte de învățare și testare. Acestea pot evidenția greșeli tipice făcute de diferite categorii de persoane care studiază limba engleză ca limbă străină la diverse nivele de performanță. Rezultatele studiilor efectuate pe Longman Learners' Corpus și Cambridge Learner Corpus pun în evidență greșeli tipice care au fost preluate de dicționarele pedagogice începând cu anii '90 (*Longman Language Activator* din 1993, *LDOCE3* și *CIDE* din 1995) și mai apoi de materialele de studiu publicate de cele două mari edituri.

Pe scurt, convergența lexicografiei pedagogice și a lingvisticii computaționale a produs o serie de dicționare pedagogice monolingve pentru străini inovative și influente, care au inspirat o nouă manieră de descriere a fenomenelor lingvistice, noi teorii de învățare și, mai presus de orice, progres lexicografic. Dicționarele pedagogice monolingve pentru străini au preluat principiile dicționarilor din perioada pre-computațională, au rafinat și au făcut mai abordabile definițiile, au descris comportamentul sintactic al cuvintelor folosind scheme tot mai transparente, au selectat cuvintele explicate în funcție de frecvență, au oferit explicații ample privind sinonimele, antonimele și hiperonimele și au abordat dificultățile de învățare prin prisma prevenirii erorilor.

Modificările cele mai recente survenite în lexicografie au dus la apariția e-lexicografiei, determinată de noile tehnologii ale informației și de expansiunea Internetului. Aceasta însumează diferite produse, de la dicționarele tipărite publicate online, cele pe CD-ROM și cele funcționale (customizate, personalizate), până la redactarea colaborativă ('crowdsourcing') a unor site-uri lexicografice (cum este *Wiktionary*). Viitorul e-lexicografiei pare a fi legat de aplicații personalizate sau chiar de sistemele de operare ale calculatoarelor.

Partea a doua a lucrării prezintă succint câteva domenii ale lingvisticii aplicate care pot oferi teme generoase de studiu pentru tinerii cercetători, precum: limbajele de specialitate, multilingvismul, predarea, testarea și evaluarea limbilor străine, pregătirea profesională inițială

și continuă a profesorilor de limbi străine, relațiile dintre lexicografie și educația lingvistică, impactul tehnologiei asupra educației lingvistice.

Cuvinte cheie: lexicografie pedagogică, dicționar pedagogic monolingv pentru străini, principii cumulative, *ISED*, lingvistică computațională, *COBUILD*, frazeologie, e-lexicografie.

Summary

The paper “Accretion: The Case of MLDs” follows the evolution of a typically English lexicographic product, the monolingual learner’s dictionary (MLD), and illustrates how pedagogical lexicography acts as an interface between theoretical and applied linguistics and the teaching of English as a foreign language. Lexicography is a field which benefits largely from the contributions of theoretical linguistics as it is the place where theories and principles necessary for language description are formulated. And as the distinction between theory and description has not been clarified yet, one can say that the first lexicographers marked the beginnings of linguistics, theoretical and applied, and accompanied those of the teaching of English as a foreign language.

While the interest of language theorists can be limited to language or social errors from the perspective of the deviations from the norm, the interest of the applied linguists and practitioners alike concentrates on language use from the perspective of recommendations, and the need for an authority able to explain what is acceptable according to standards, norms and conventions, and to provide examples. However, recommendations have theoretical support, based on considerations of language regularization and the factors that influence it. Both the prescriptive and the proscriptive views are linked to standards, and combine in the modern pedagogical dictionaries by integrating grammar and lexicography, making language theory, description and use converge. Therefore, dictionaries are given the respect due to the authorities which contribute to language standardization. Although in the 20th century both dictionaries and grammars became mainly descriptive, they could never ignore norms or standards and preserved the role of authorities, as shown by the ever-growing market for dictionaries, grammars, spelling and punctuation guides.

Pedagogical lexicography is an accretive activity with a history of thousands of years which began with the proto-dictionaries carved in clay used to train the Babylonian scribes. British pedagogical lexicography started in the 16th century with the word lists published by the practitioners who were trying to standardize spelling and explain ‘hard’ words. As English has expanded in time and won the status of language of world communication, its study has become a global activity, and the market for pedagogical materials has increased constantly. The number and variety of such products point to the leading role of present-day lexicography in producing systematic inventories of information and its dissemination. Since 1942, the four

great British publishers – Longman, Oxford, Cambridge and Collins – have issued regularly monolingual pedagogical dictionaries for foreign learners with various degrees of proficiency. More recently, they have been joined by American, Canadian, and Australian publishers.

The paper describes the special status of the monolingual learner's dictionary among the other pedagogical dictionaries and offers an account of the slow elaboration and accretion of lexicographic conventions, starting with the dictionaries of the first British lexicographers, Robert Cawdrey, John Bullokar, Nathan Bailey and Samuel Johnson, and evinces the close connections between theory, application and practice.

The chapter that deals with the contribution of the 'Vocabulary Control Movement', Harold Palmer, Michael West and A. S. Hornby, analyzes the first MLD, the *Idiomatic and Syntactic English Dictionary (ISED)* of 1942 and its subsequent editions: *The Advanced Learner's Dictionary*, *The Oxford Advanced Learner's Dictionary of Current English*, *The Advanced Learner's Dictionary of Current English* and *An Advanced Learner's Dictionary of English*. The Movement inspired the principle of the close connection between grammar structures and lexis and that of identifying a minimum vocabulary based on word use frequency. Palmer and Hornby's interest in pedagogical grammar led to the inclusion in *ISED* of the first verb patterns, transforming the dictionary into the first illustration of the basic pedagogical lexicographic principles: word selection based on frequency and usefulness, inclusion of standard pronunciation using phonetic transcription codes, the use of simple and direct definitions, the illustration of meanings by contextualized examples, the presence of explanations on synonyms and grammatical information including verb patterns, collocations, register, and the presence of pictures and appendices. Another linguist who influenced the evolution of the MLD was Michael West by the publication of the *General Service List of English Words (GSL)*, a document that contains the principles of word selection based on frequency, structural value, universality, semantic coverage, flexibility and style. In brief, the accretion of the characteristics of the MLDs in the pre-computer era made them unique and contributed to the gradual formalization of a view of the lexicon and grammar which is still in use today.

The second period in the evolution of MLDs is marked by the collaboration between lexicography and corpus linguistics, whose first fruits were the *Oxford Advanced Learner's Dictionary of Contemporary English*, *Longman Dictionary of Contemporary English* and, above all, *COBUILD*, all published in 1987. Corpus linguistics offered a new perspective on language, a new way of describing it and new theories of foreign language learning and acquisition, all based on data and concepts. With the replacement of the data used for analysis,

from intuitive to natural, authentic ones, corpus studies have redirected linguistics and foreign language teaching from language knowledge to language use. The object of linguistic description has become the recorded spoken or written texts, authentic and appropriate to the contexts in which they were produced. The empirical analysis of natural data is not, however, an innovation due to corpus linguistics. It was used by Samuel Johnson for his *Dictionary of the English Language* of 1755, Noah Webster for his *American Dictionary of the English Language* of 1828 and *The Oxford English Dictionary* of 1928, all based on manually collected corpora. However, the larger databases of the '80s and the new data processing techniques started to show discrepancies between real and introspection data analyses and made possible more accurate and detailed descriptions of structures, regularities and tendencies that had not been detected before. Moreover, the differences between the spoken and the written modes could be documented more convincingly (e.g., Biber *et al.* 1999; Carter and McCarthy 2006).

The year 1987 marked a turning point in British lexicography by the publication of John Sinclair's *COBUILD (Collins COBUILD English Language Dictionary)*. This innovative dictionary is based exclusively on the analysis of a massive corpus, and uses authentic examples and full-sentence definitions only. The definienda are given in their typical occurrences, illustrating the meaning – use connection, and the presence of typical collocations illustrate the principle according to which each difference in meaning is associated to a formal one. Actually, *COBUILD* was the first dictionary to demonstrate that the co-selected structures affect word meaning, in other words, that word meaning is as much a property of word sequences as it is one of the single word. Sinclair followed the principles formulated by the 'Vocabulary Control Movement', combined with the advantages offered by the progress of information technology. With *COBUILD*, the principle of frequency becomes essential in dictionary compiling, and frequency, signalled by various symbols, features besides semantic, pragmatic and grammatical information. Another lexicographic innovation are the definitions consisting in full sentences, formulated in familiar terms and similar to examples. Despite criticism, *COBUILD* remains one of the most influential pedagogical dictionaries of all times.

After *COBUILD*, lexicographers reconsidered the idea that a word in isolation can have several meanings and that dictionary definitions are precise and mutually exclusive. Patrick Hanks showed that a word has only meaning *potentials* and that senses derive from the structures in which it is used. Together with Pustejovsky, Hanks also showed that the meanings of a word are not mutually exclusive and that there are no certain criteria for distinguishing one meaning from another and, that meaning is vague, imprecise and probabilistic. Language itself is a probabilistic and preferential system in which contexts correlate with statistically relevant

patterns and collocations. This kind of preferences and probabilities can be presented only in a new type of dictionary, a pattern-based one. These new theories have led to new principles of dictionary writing which announce a new type of dictionary based on meaning – pattern associations, and possibly to a new period in the development of lexicography.

In the meantime, a new type of corpus, the ‘computer learner corpus’, which contains texts created by students in learning and testing contexts, has been created. Such corpora can show typical errors made by different types of with various levels of performance. The results of the analyses performed on Longman Learners’ Corpus and Cambridge Learner Corpus have been incorporated in the pedagogical dictionaries since the ‘90s (*Longman Language Activator* of 1993, *LDOCE3* and *CIDE* of 1995) and in the study materials published by the two great publishers.

In brief, the convergence of pedagogical lexicography and corpus linguistics has produced several series of innovative and influential MLDs which have inspired a new way of describing linguistic phenomena and lexicographic progress. MLDs have taken over the principles of pre-corpus lexicography, refined the definitions and made them more friendly, described the syntactic behaviour of words using ever more transparent patterns, selected definienda according to their frequency, offered ample explanations concerning meaning relationships, and approached learning difficulties via typical errors.

The most recent change in pedagogical lexicography is the advent of e-lexicography, which involves the contribution of the new information technology and the expansion of the Internet. e-Lexicography comprises various products from printed dictionaries published online, to those on CD-ROM and the functional (customized, personalized) ones, to the crowdsourcing of lexicographic sites (*e.g. Wiktionary*). The future of e-lexicography seems to be linked to personalized applications and even to the computer operating systems.

The second part of the paper is a succinct introduction of several areas of applied linguistics which can offer generous topics of study to young researchers: special languages, multilingualism, the teaching, testing and assessment of foreign languages, initial and ongoing teacher training, the relationships between lexicography and language education, the impact of technology on language education.

Key words: pedagogical lexicography, monolingual learner’s dictionary, cumulative principles, *ISED*, corpus linguistics, *COBUILD*, pattern, e-lexicography.