The Oxford-DZS Comprehensive English-Slovenian Dictionary

Mojca Šorli

Trojina, Zavod za uporabno slovenistiko Partizanska cesta 5, SI – 4220 Škofja Loka Slovenia mojca.sorli@guest.arnes.si

Coauthors: Katarina Grabnar, katja.grabnar@guest.arnes.si; Simon Krek, simon.krek@guest.arnes.si; Tomato Košir, tomato@tomato.kosir.com

Abstract

The paper presents the most comprehensive English-Slovenian dictionary to date. As a corpus-based work with approximately 120,000 entries and an in-depth treatment of the material, the dictionary sets a new standard in Slovenian bilingual lexicography. Designed as a monodirectional bilingual tool, this desk-size dictionary is characterised by a number of specific solutions which might be of interest to a broader lexicographic community. The first section deals with the entry headword list, the second with the rendering of metalanguage, and then follows a section on the grammatical and semantic treatment of the entry headword, with the sub-sections elaborating on sense categories, translation equivalents, and illustrative examples. The paper concludes with a section on the dictionary's layout and format.

1 Introduction: A Monodirectional Bilingual Dictionary

The Oxford-DZS Comprehensive English-Slovenian Dictionary (henceforth ODCESD) is special on account of two editorial decisions. The first was not to design a revision of the existing English-Slovenian dictionary, but to compile from scratch an ambitious bilingual work, which, in terms of macrostructure and microstructure – within the scope of a standard desk-size dictionary – granted its compilers virtually unlimited space to work with. This is reflected both in the extensive and in-depth treatment of the lexicographic material as well as in its opulent layout. The second decision, related to and ensuing from the first, was to design a dictionary as a decoding tool for Slovenian speakers only, rather than for English speakers who want to produce (encode) Slovenian texts. The result is a dictionary containing 120,000 entries which addresses the needs of only a quarter of the potential four-type² users of, for

¹ Veliki angleško-slovenski slovar. A Comprehensive English-Slovenian Dictionary. (2004/1978).

² The French-English section is designed for the decoding processes in English speakers and the encoding processes in French speakers, whereas the English-French section is designed for the reverse processes. The abbreviation OHFD3 consistently refers to the English-French section only.

example, the bi-directional bilingual *Oxford-Hachette French Dictionary* (henceforth OHFD3),³ but does so truly comprehensively. This practical decision bears influence on a number of specific solutions regarding the entry headword list, metalanguage, the treatment of sense categories, and the selection of illustrative material. In quite a few elements of lexicographic description, the monodirectional bilingual ODCESD appears to be further from bi-directional bilingual dictionaries than from monolingual dictionaries (inclusion of archaic or rare entry headwords/meanings, illustrative material, phraseology, and metalanguage). Apart from the basic principle of providing translations, there is not much more resemblance between ODCESD and OHFD3 than between ODCESD and an EFL such as LDOCE4. The category which is entirely ODCESD specific is sense distribution.

2. The Entry Headword List

2.1 The scope of entry headwords

ODCESD includes not only the current lexicon of English, but also rarely used words, e.g. encipher, en clair, farthingale, huggable, ice chest, and even archaic words, e.g. egad, holla, illation; it contains a great number of technical terms, e.g. endemism, fibrosis, fermion, idiotype, imagist, and also covers abbreviations and proper nouns. Here too, like in all of the other listed categories, it goes beyond the scope of an average-sized bilingual dictionary, as a random comparison between OHFD3 and ODCESD confirms. The OHFD3 entry headword list from i to id contains seven abbreviations, i.e. IA, IAAF, IAEA, ib, IBA, IBDR, ICPO, whereas that of ODCESD includes twelve more abbreviations, i.e. IAA, IATA, IBS, i/c, IC, ICAO, ICBM, ICC, ICE, ICFTU, I.Chem.E., ICRP. A similar comparison of proper nouns on a slightly larger section of both entry headword lists, i.e. from i to if, shows that OHFD3 includes six proper nouns, i.e. Iberia, Iberian Peninsula, Icarus, Icelande, Icelander, Idaho, while ODCESD adds Iapetus, I Ching, and Idomeneus to the list, but leaves out Idaho, giving a total of eight proper nouns.

As a decoding tool, however, ODCESD does not include affixes. The meaning of prefixes and suffixes in isolation would only be of use to speakers/writers who wanted to form new words, while the users of ODCESD look up translation equivalents of whole words (made up of a root and an affix). Thus, for example, the prefix anti- is not an entry headword, but numerous words formed with this prefix are, e.g. antiabortion, anticlimax, anti-inflation, antisocial, etc.

2.2 Derivatives

In accordance with ODCESD's comprehensiveness, the list of entry headwords is expanded with run-on derivatives from *The New Oxford Dictionary of English* (henceforth NODE) (only run-ons of the primitives from ODCESD were taken into account). A derivative is included if it is lexicographically relevant, one criterion being a sufficient number of hits in the *British National Corpus*, e.g. crushable (6), crustal (114), curatorial (58); derivatives with-limited or zero occurrence are omitted, e.g. crumbliness (0), crumply (2), crustily (0).

³ OHFD3's English-side material was used as the basis for ODCESD's English side.

⁴ BNC: http://www.natcorp.ox.ac.uk/.

3 Metalanguage

While in ODCESD all metalinguistic information, namely register and field labels, regional labels, sense indicators, and grammatical data, is in the target language (Slovenian) only, in OHFD3 metalinguistic information is both in the source and target languages, depending on which user it addresses. The target language in ODCESD is not described metalinguistically, as it is presumed that the user is a native speaker of Slovenian. French register labels⁵ in OHFD3 refer to the English material (entry headwords and illustrative examples),

bite (batt) I not (mouths) bouches f. in one - en use beachda; to have on take a:of ath prepare una beachte de qch; to take . - out of seth to talre, up, treu dans qob; that will take a big - out of our budget/profits cela va taire un grand trou dans notre budgetmango beneficialre: 20 (susci) mercean w (a manger); to have a (to est) manger un morceau; to have or grab . a . quick . — (to ess) manger . un manceau en vitesse: 3 de (intest, keen edge) (of wind, cold) moreure f; (of food) piquest ni; (al agunes, carbonasca, ciyla, film) mordani his speechiffing has M. SOIL discouration a do merdant: 4 (from Insect - piqure / d'insecte: 5 Fisig toucho f, to have a - it avoir une couche: ip trouver amatour; the house is up for sale but we haven't had any es yet la malson est en vento mais nous n'avons pas encore trouvé amatour: 6 Out occlusion & ... IL sir (pre bill, pp bitton) [person, extent] mordine: [trisen] piquer; to - eth in two conper lond en deux d'un comp de dent; to - one's nails se rouger les onglis..... III ni (prés bit, pp bitten) 1 (take effect) measure, colley, rule, new rates, strike, shortand se faire sentir: 2 Fisho [Ath] mordre. ment - your light no va pas to mangard! to - one's lip to mordre its levres; to - the hand that feeds you creeker dans to coupe; the bites bit tel est juis qui croyali prendre; to be bitten by the DIV/health food bugs attroper la virus? du bricologo, de la diciétique: bullet. bite back: - back (eth) ravaler (roof common, reply!

Bite into: ~ into [sta] it merting dans Liture, soustwich errij: 5g (slied) avetr un elliri term [actioning Anonces]. a bite off - off (eth), - (eth) off min cher [geli] O'on coup do deza. a bite on: - on (ath) mordre sur-bite through: - through (ath) [pryon colon] percer [qch] avec see dents. likewise grammatical information is aimed primarily at the French users and is therefore in the target language (OHFD3: bite B vtr (prét bit, pp bitten)). Another major distinction between OHFD3 and OD-CESD is the use of field labels. The former employs them freely to indicate semantic fields of the entry headword (or one of its senses) rather than marking its specialist term status. ODCESD, on the contrary, uses field labels to mark the entry headword (or one of its senses) as a terminological unit. In ODCESD, all of the above listed categories, with the exception of sense indicators and the label figurative (fig.), refer strictly to the source language (English) (bite II). While OHFD3 renders sense indicators mostly as synonyms to distinguish particular senses of the entry headword (bite A 1, 2, 3 etc.), ODCESD employs them, quite differently, to determine the scope of the translation equivalent(s), or to disambiguate the Slovenian translation(s) (bite I 5, 6, 7). OHFD3 uses collocates extensively to facilitate the encoding process (bite B, C 1, 2), whereas in ODCESD the most typical collocates are shown in examples, as part of the context (bite I, 2: a dog bite). Their role, therefore, is not to point to the right translation, but to suggest the translation of the most typical collocations. Grammatical information in ODCESD, i.e. phonetic and morphological information (parts of speech, irregular plurals in nouns, forms of comparison in adjectives, and verb forms) is provided for the source language (English) only.

Figure 1. Bite - OHFD3

⁵ In opposition to the English register labeling, the French register labeling is done by symbols.

town a court plain (demonstrate par terray blathing extensibility reported the round by term to the bone may a rue of they will make and best back toward as bods casacili do bessi (ary wined bits one forces between in relative accord il prid. क्षा है। record system was be record to extension क्षेत्रक्षां क्षेत्र । १ - शतक्षां क्षा क्षेत्रक्षां क्षा १ रजस्यों क्षा Beirrkui o symentic g aon, esperinajeriri, ofosii princeri III ricpyrib, ging, P birt, FI birticoni a wegori selit, Expris piri duce isi olag birt, il spinov so granti bir selita, il will beri seleturi bir rath agricul v ca, luigi bei gannosi bastolicoj bitzko (Tagis polsk rom spanja, dietern, budeben, ffictie, zkon anteren, ikonda bellen besta a machymaned goda rela ja ja bazty the same hit; into hits ship over a counte catrix v transi, the catif for it was my borson water ermont delesi primento (no tertitatio altori di californi con lega descalitação (dun ess rat 's regal do bond a r-adjeant; materia, eggreciti the single payer the strength and the inches a bierter de na rode roje bilo podlo de mindo kultim kas on milit magaza otjeta ktole (ko ost hin hun a v lakati sapristina judia kwa ti. delevati bearing his insolution directly I SEEL r sgrtif)z(, zaloga) conjugative of the confliction o bier (briefi) un. Linkalede, Ugie nosnedaniki ukarakto pokutuli zobe i performs the state person and a second secon ran (but la com late repo e res do marabrat acad e caro (but will take a baj biso sa rei than dog is a historyckel per gritte i this en beden income a la crematida ed bullon bias of consussion de del papero inentitik mendifikan is berlikti pikip pinjah kecka preperand bisat a suadn rik withinkers hi see wint i portion proces to distante solution 4 saughal, aujeut, aufeat; audeat, and one persons in gave at the agreem has the fire was overeid a measure bear dear a most on conduct and homeyor yours, projection, making where you'd him least to be worthly print 4 provings, contract, that fivest, planning, led know, that morning dimor the what has been that morning dimor e the biter bit we, en Sie et eggiske wike za drego bisoelerië i funtamië pid. 1-uilt nagrifije rata è roti the said is scaring so the schemath the s enter libiar landoù mensko essiba: eterino fidera de rodora comente persiste bis bevisqui duni dei dei notorgi es, della je parcural, se il la rota la glabadan e seca g a l'apar o bodendi jeri jeri je vejadani venente ne, pori jeri je verginem cuncipal Raggradd la l'oprijeri فوضيونيه bin-that damin of discharge gality wild z a-drobers, kraźck: obwładi ko Mars sked merupage i komi krab i komina darid i han-dadi panikan daridijisi padikad Didigaria Alui Varuzi sam erter cutto britt: Mix aponch han bike njegov the etectude not but halou or primare never into extrinct wine some bits to never enak to charging the imporporati han the sear with the reck! pa, do ho Perjo glados kolj calasso i Bils GEAL HOMEN Bidroja (orbitaje izve za deždo poer pro grant and represent types of the control o in the cety merchan the fide bits conduir kladi (mir) chain) i be palendoja (ba kasw i i wa a manik: bur be ki asyway welel je. bates June I ame. ease a periolist (MADASA) e fake ili isi chimerirei in eo les mai apodés je bed, propried jesa de es cala je nancapes provar the diseas. t a osser, bad, strupen, resek the bang wind thilled on to the bone 6 e let actions àctions/escopir, plicationest, pelot disse remoss diction lette les molecules de decig a second blue as the charry sa. nos prikinos erapen som me is provinci do bend a raqied (ibe, pilker, jedick, comer, menapen he carefuelt for a second bits at the cherry wall ach: delt productioners a no be bittern by the bug/crasss proteil obsection a time achiese marketid entail his bid ag some a lagren cake liberes i calcion il like wine with his cul bana fine biling, racement to choose the generages: the most belong racine claudes (softer andre gargelingsal), ki gener 7 e japo e boloskij oprijene: "V avejama avaja j migrabitov, prijene; vrsa, mest to bonomethe bas of the wheels intelligat he was bissen by the crowd tage regards to these bising insects and played insets biclagily (furnels) ond. i iver bedes "hi bi a midir l e to bite the big one sea us raci S of upo propoliticated access, acciden zo, hudo, strupceo, miko pojedkarje pojedkarje poseopniem jemanie sido modera sentanie udar seco korean ke print as blac the hand that feeds you the sirkers as his hage ordered to be possible analysis was bringly irkizatier refreistanza na istmato i incasinali eta (berz pouz bej nateniati kopa piarril az kopa na bizo onno i ilp ngrimatien v jezik; ermor ekus cultuk hibidrum rezudu. 2 - Enjedijiro, pikro, jedda, osta, erticus Batto dinines o binario mai a bingga arcaka akgaragata marakajaka mituskin binario (hiterap) kat. is seen and the is arm, segment growing and processing of the control of the contr andstation De pens off more than one can
 they lottle built more than one lottle
 in a way and it more than one i ser. • bina dita Of the few bases on Oracle (SOP 10 - 59-1 - capitati v bimi siki e to bite (en) the fullet roccutariti robs i in this side in inn presentation ha poste gaing to how to his dis balks 2 **-** pa**c**ina i was beneate a make bata ne k pilita. James o grissis, fazzi, sancisi a blir areament houri biarra pieces. and her kim much but almid orbe in sc o lithe won't bite you! roa all negated etos) ili i appeal kai saprati 4 - ja farani, pijarij anpeal, paligani - mailteatstranska vlega bir player ram. ni negrind + to put the tite on the an initit denied lega. Strawijchi shadoni: bite back julij i- zadašii, shadiniii, polostii zadažai, shadiniii zase whitees blace the threat with picker reserbs bit-rane care.

sac. + bitrare persons podution
bitraream (bundars) care
task. + bitrarias g = magalest, majeret, majeret, majeret racional, racional ratgions, institutes the artificione durante la librar estitus herius de algue, o habathi aprificial sa; to problema mordial (2021) this prificial these synta have de thindry our di mag an tey read or unume to problemation artificial ships phonois i the change with him when which whose is the change with him when which whose s Minuream' - orementa htracy aics ste bû bech der einn reertrafil is seter i a Korner Christo (Deschild): aparers has biorgania ha bit luich the remark proposite ja tadrād biosy / beta/ prkl.

man na emics, magicen
bioten/biren/part. del.

g* birs the give dries spons be wished for delter throughput was see before the could arlipale on praedi y e anniediti se, annenii frek anjerti se; rarini bite is been synchrafe of je calls, in presiden mini la e krah

Figure 2. Bite – ODCESD

4 Grammatical and Semantic Treatment of the Entry Headword

4.1 Sense categories

In the typologically specific ODCESD, the organisation of the senses is translation-based. In its English-French section, OHFD3 provides, on the one hand, (monolingual) sense

distribution for an English speaker who is looking for a particular sense of the entry headword, and, on the other, the translation equivalent(s) sought by a French speaker. Once the former has located the right sense, (s)he is assisted in selecting the appropriate translation by English collocates listed in groups according to the French translations. ODCESD, as a monodirectional bilingual dictionary, takes quite a different approach: the emphasis is on joining all the senses which generate the same translation equivalent(s). The repetition of translation equivalents within the same entry under more than one numbered sense is avoided. Generally speaking, compared to bi-directional bilingual and monolingual dictionaries, in ODCESD there are fewer sense categories in an entry. However, old-fashioned, archaic, and rare senses, relevant in the decoding processes, are also included.

4.2 Translation equivalents (synonymy and ordering of translations)

OHFD3 often lists only one translation equivalent where in ODCESD the translations listed under each numbered sense are an exhaustive set of semantic and stylistic equivalents. Synonymous translations are exceptional in OHFD3, listed only when the two or more equivalents are completely interchangeable. Separated by a comma, close synonyms in ODCESD are formally treated as full synonyms and are allowed to mutually determine each other semantically, regardless of whether their synonymy is valid from a monolingual point of view (bite I 5, II 5). To be treated as such, close synonyms must be interchangeable in more than 50% of the examined contexts. Those equivalents with a higher degree of specific collocate distribution are separated by a semicolon. They do not require a separate numbered sense, but are, at the same time, not considered close synonyms due to differences either in intensity of meaning or connotation.

Each sense category offers a range of possible translations. First listed are the translations which are semantically closest to the entry headword and which are applicable to the broadest range of contexts. These are followed by the translations which are useful in more specific contexts (bite I 4, 6; II 7). Typically, the list of translation equivalents is followed by an exhaustive set of illustrative examples, depending, of course, on the level of semantic complexity of the entry.

4.3 Illustrative examples

Illustrative examples are selected and treated strictly as instances of natural text. Unlike the provided translation equivalents, which aspire to universal solutions, the illustrative material conveys both the uniqueness and the particularity of an individual speaker's choice.

Grammatical structures and other structural examples, typically highlighted in bi-directional bilingual dictionaries such as OHFD3, or in EFL⁶ dictionaries, but less so in monolingual dictionaries such as NODE, are relatively limited in number. The emphasis is placed on typical collocations and contrastively relevant examples, the latter, in particular, showing the role of the context in the decoding processes. The highest in the ordering of the provided ex-

⁶ Monolingual dictionaries for foreign learners of English, such as LDOCE4, or CCELD2.

amples are those which can be translated into the target language directly, by one of the listed translation equivalents (bite I 2). These are followed by examples in which the context noticeably influences the translation, and/or no listed equivalent corresponds to the entry headword on a one-to-one basis. Underlining its decoding purposes, a distinguishing feature of the dictionary is an overt preference for generously contextualised examples, including grammatical collocations (bite I 4; II 6). Nevertheless, when the translation into the target language is relatively unproblematic and generally applicable, a structural example is preferable for its conciseness (bite: to give sb a bite; to have a quick bite to eat; to bite sth in two).

Most frequently, the examples listed for contrastive purposes show the required syntactic transformations, such as a noun to verb or a noun to adjective shift, or the so-called zero translation (bite I 4). Thus in (bite I 1, 2), instead of selecting and translating a short collocation a mosquito bite simply as pik komarja, lit. "the bite of a mosquito", ODCESD provides a contextualised example his face was covered in mosquito bites, which commands a noun to adjective/past participle shift in obraz je imel ves popikan od komarjev, lit. "he had his face all bitten by mosquitos", offering a context-dependent translation. In (bite I 4), this issue has bite is rendered into Slovenian with a highly idiomatic to vprašanje buri duhove, lit. "this issue stirs the spirits". At the very end are listed specific, sometimes field-labelled contexts. Idiomaticity of translation, guaranteed by the rigorous consideration of the Slovenian corpus, is pursued to the point that often none of the translations employs the listed Slovenian equivalent(s) of the entry headword. This occurs primarily in two cases: with headwords whose translation is less context-sensitive, typically nouns (bite I 3), and with headwords whose translation is highly idiomatic, often verbs.

5 Layout and format

Most contemporary dictionaries were originally typeset in lead, which placed many restrictions on typesetting, and new technologies, including computer typesetting, seem to have adopted these old solutions uncritically. Consequently, the designer of ODCESD's layout re-examined the traditional approach and came up with a few unconventional solutions, most notably:

A generally 'spacious' layout is based on the key premise that form should derive from function. The general way of proceeding is from micro- to macrostructure, even though the user's approach is the reverse. A broad page format allows for a considerable shift of the body text from the centre, i.e. from an area that loses functionality due to the large number of bound pages.

A wide range of family typeface designed for lexicographic use was selected. This range of letters is characterised by good readability in even the smallest font size and highly suitable for printing on coated paper, by a good height of minuscule letters, and by a wide range of bold typeface used consistently for the italic variants, small caps and old style figures. It also provides pronunciation symbols. Size variation in typographies is limited due to the need for a rational use of space.

⁷ The Reference Corpus of Slovenian FIDA: www.fida.net.

References

A. Dictionaries

Butterfield, J. (ed.) (2003), Collins English Dictionary. (Sixth Edition.) Glasgow, Collins.

Grad, A., Škerlj, B., Vitorovič, N. (eds.) (2004), Veliki angleško-slovenski slovar (First edition 1978), Liubliana. DZS.

Pearsall, J., Trumble, B. (eds.) (1995), Oxford Reference English Dictionary, Oxford, New York, Oxford University Press.

Sholze-Studenrecht, W., Clark, M., Sykes, J. B., Thyen, O. (eds.) (1996), Oxford-Duden English-German Dictionary, (Second Edition). Oxford, Oxford University Press.

Dictionaries (with abbreviations):

Corréard, M-H., and Grundy, V. (eds.) (2001), Oxford-Hachette French Dictionary. (Third Edition.) Oxford, Oxford University Press. (OHFD3)

Duval, A., and Back, M., (eds.) (1998), The Unabridged Collins-Robert English-French, Dictionary. (Fifth Edition.) Birmingham, HarperCollins. (CREFD5)

Krek, S. (ed.) (2004), The Oxford-DZS Comprehensive English-Slovenian Dictionary. Vol. 1 (A-K), (First Edition.) Ljubljana, DZS. (ODCESD)

Pearsall, J. (ed.) (1998), The New Oxford Dictionary of English. (First Edition.) Oxford, Oxford, University Press. (NODE)

Sinclair, J. (ed.) (1995), Collins COBUILD English Language Dictionary. (Second Edition.), London, HarperCollins. (CCELD2)

Summers, D. (ed.) (2003), Longman Dictionary of Contemporary English. (Fourth Edition.), Harlow, Longman. (LDOCE4)