Reimagining the Dictionary, or Why Lexicography Needs Digital Humanities

Tasovac, Toma

ttasovac@humanistika.org

Center for Digtial Humanities (Belgrade), Serbia

The promise of eLexicography stems not only from the transformation of the production medium, but also from the technological feasibility of representing linguistic complexity. Even though modern lexicography is unimaginable without computer technology (Hockey, 2000a; Knowles, 1989; Meijs, 1992), the sheer use of computers in producing a dictionary or delivering it electronically does not automatically transform a dictionary from "a simple artefact" to a "more complex lexical architecture," to use Sinclair's (2000) formulations.

Calling dictionaries "simple artefacts" is itself a rhetorical oversimplification: there is certainly nothing simple about a dictionary — whether we look at it as a material object, cultural product or a model of language. Yet the overall structure of dictionaries as extended word lists has not changed in centuries (Hausmann et al., 1989; Fontenelle, 2008; Atkins and Rundell, 2008). Admittedly, a great deal of factual information is packed into a prototypical lexicographic entry, but a defined term often remains in isolation and insufficiently connected or embedded into the language system as a whole. This is what Miller refers to as the "woeful incompleteness" (Miller at al.) of a traditional dictionary entry, and what Shvedova sees as its "paradoxical nature" — dictionary entries tend to be "lexicocentric" while language itself is "class-centric" (Шведова, 1988).

Furthermore, the advances in digital humanities, textual studies and postmodern literary theory do not seem to have had a profound effect on the way we theorize or produce dictionaries. Surely, many important lexicographic projects have been digitalized and gone online; web-portals increasingly offer cumulative searches across different dictionaries; and eLexicography is a thriving field (Lemberg et al., 2001; Hockey, 2000a; de Schryver; Hass, 2005; Nielsen, 2009; Rundell, 2009; Hass, 2005), yet dictionaries — often commercial enterprises which are guided by predominantly economic concerns — remain by far and large discrete objects: no more and no less than digitalized versions of stable, print editions. We still consult dictionaries by going to a particular web site. Dictionaries do not come to us.

The time is ripe to ask — both in theoretical and practical terms — a new set of questions: how has the electronic text changed our notion of what a dictionary is (and ought to be); how have the methods of digital humanities and the advances made in digital libraries altered our idea of what a dictionary can (and should) do? And, finally, where do we go from here?

The dictionary is a kind of text. In print culture, the dictionary, like every other text, had its material and semantic dimension. The semantic dimension was represented on its visible surface, whereas its depth was in the mind of the reader, or what Eco refers to as the "encyclopedia of the reader." (Eco et al., 1992; Eco, 1979). Yet if we — as we should — start thinking of the dictionary as a kind of electronic text, the way Kathrine Hayles and others have done for electronic literature, we will have no choice but to strip the dictionary of its finality and its "object-ness" and see in it, instead, only one possible manifestation of the database in which it is stored (Hayles, 2003; Hayles, 2006; Folsom, 2007). A digital text can be not only edited, transformed, cut and pasted — as part of our computational textual kinetics — but is always part of other activities: search, downloading, surfing. In other words, an electronic text is unimaginable without its context (Aarseth, 1997; DeRose et al., 1990; Hockey, 2000b).

The dictionary, then, should be seen as a kind of semantic potential that can be realized through its use. But in order to truly fulfill this potential, the dictionary needs to be embedded in the digital flow of our textual production and reception. That is why we cannot think of dictionaries any more without thinking about digital libraries and the status

which electronic texts have in them (Andrews and Law, 2004; Candela et al., 2007; Kruk and McDaniel, 2009; Maness, 2006; Miller, 2005; Novotny, 2006). To be truly useful for any kind of textual studies, the digital library must "explode" the text (by providing full-content searchability, concordances and indexes, metadata, hyperlinks, critical markup etc.) instead of "freezing" it as an image, which, albeit digital, is computationally neither intelligible nor modifiable as text. In smart digital libraries, a text should not only be an object but a service; not a static entity but an interactive method (Tasovac, forthcoming). The text should be computationally exploitable so that it can be sampled and used, not simply reproduced in its entirety. This kind of atomic approach to textuality poses a host of challenges (legal, ethical, technical and intellectual, to name just a few), but it opens up the possibility of creative engagement with the digital text in literary studies (text mining, statistical text comparison, data visualization, hypertextual systems etc.).

The consequence of this "explosive" nature of the electronic text is of paramount importance for eLexcicography and the reformulation of the dictionary not as an object, but a service. We should start thinking of and building dictionaries as fully embeddable modules in digital libraries, or, to put it differently, build digital libraries which integrate dictionaries as part of their fundamental infrastructure and allow an ever-expandable process of associating words in an electronic text with an equally changeable record in a textual database. The changeability of the dictionary entry will, in turn, defer *ad infinitum* the notion of a particular dictionary edition — other than as temporary snapshot of the database. The dictionary as an evolving process will be in a permanent beta state.

The future of electronic dictionaries undoubtedly lies in their detachability from physical media (CD, DVD, desktop applications) and static locations (web portals). If we think of the dictionary as a service with an API [1] that can be called from any Web page, we can actually start thinking about any (electronic) text as a direct entry point to the dictionary. If every word in a digital library is a link to a particular entry in the dictionary, electronic textuality as such

becomes an extension of lexicography: the text begins to contain the dictionary in the same way that the dictionary contains the text.

The Center for Digital Humanities (Belgrade, Serbia) is putting these theoretical considerations into practice while working on its flagship *Transpoetika Project* (Tasovac, 2009). *Transpoetika* (see Figure 1) is a collaborative, class-centric, bilingualized Serbian-English learner's dictionary based on the architecturally complex, machine-readable semantic network of the Princeton Wordnet (Fellbaum, 1998; Vossen, 1998; Stamou et al., 2002; Tufis et al., 2004). It is part of a scalable, web-based, digital framework for editing and publishing annotated, fully-glossed study editions of literary works in the Serbian language, primarily aimed and students of Serbian as a second or inherited language.

Transpoetika has been designed to be deployed as a web service and therefore linked from and applied to a variety of textual sources online. Portions of the project, such as the Serbian Morpho-Syntactic Database (SMS) already function as a web service internally and will also be made public and free once the sufficient funding for the project has been secured. *Transpoetika* can also interact with other web services: by using Flickr as a source of illustrations, and Twitter as a source of "live quotes" in the entries, the *Transpoetika* Dictionary explores the role of serendipity in a lexicographic text.

The overarching goal of the Belgrade Center for Digital Humanities (CDHN) is to produce a pluggable, service-based, meta-lexicographic platform for the Serbian language, which will interact with various Web-based digital libraries, and contain not only our own bilingualized Serbian Wordnet, but also historical Serbian dictionaries that the CDHN is digitalizing, such as, for instance, the classic Serbian-German-Latin Dictionary by Vuk Stefanović-Karadžić (1818 and 1852). The platform could, in theory, be extended to include and consolidate a number of other, more specialized, lexicons. This is, in any case, the general direction we would like to take.

I would like to conclude with a *hysteron-proteron*, which, in Samuel Johnson's Dictionary of the English language was defined

as "a rhetorical figure: when that is last said, which was first done." From the very beginning of this paper, I spoke of the dictionary, which every careful reader would have marked as a serious lexicographic faux-pax. There is and never was such a thing as a singular and uniquely authoritative source of information about words and their meanings. There is no such thing as the (Platonic, ideal) dictionary but rather a myriad manifestations of its imagined hypertextual prototype. I believe, nonetheless, that we should, in the digital age and with the ongoing developments of the digital humanities, reclaim the dusty notion of the dictionary and boldly, though not without self-irony, keep trying to imagine what that "thing" — the dictionary — could be. If only with the goal of making it — in its traditional, leather-bound, sense — completely obsolete.

Figure 1

References

Aarseth, E.J. (1997). *Cybertext: Perspectives on Ergodic Literature*. Baltimore, MD: The Johns Hopkins University Press.

Andrews, J., Law, D.G. (eds.) (2004). *Digital Libraries: Policy, Planning, and Practice*. Aldershot, Hants, England; Burlington, VT: Ashgate.

Atkins, B.T.S., Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford; New York: Oxford University Press.

Candela, L., et al. (2007). 'The DELOS Digital Library Reference Model: Foundations for Digital Libraries'. Vesrion 0.98. http://www.delos.info/index.php?option=com_content&task=view&id=345.

de Schryver, G.-M. (2003). 'Lexicographer's Dreams in the Electronic-Dictionary Age'. *International Journal of Lexicography*. 143-199.

DeRose, S., et al. (1990). 'What Is Text, Really?'. *Journal of Computing in Higher Education*. **1**: 3-26.

Eco, U. (1979). *The Role of the Reader: Explorations in the Semiotics of Texts*. Bloomington: Indiana University Press.

Eco, U., et al. (1992). *Interpretation and Overinterpretation*. Cambridge; New York: Cambridge University Press.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Cambridge, Mass.: MIT Press.

Folsom, E. (2007). 'Database as Genre: The Epic Transformation of Archives'. *PMLA*. **122**: 1571-1579.

Fontenelle, T. (2008). *Practical Lexicography: A Reader*. Oxford; New York: Oxford University Press.

Hass, U. (ed.) (2005). *Grundfragen der elektronischen Lexikographie : Elexiko, das Online-Informationssystem zum deutschen Wortschatz*. Berlin; New York: W. de Gruyter.

Hausmann, F.J., Reichmann, O., Wiegand, H.E. (eds.) (1989). *Wörterbücher: ein internationales Handbuch zur Lexikographie*. Berlin; New York: W. de Gruyter.

Hayles, N.K. (2003). 'Deeper into the Machine: The Future of Electronic Literature'. *Culture Machine*. **5**.

Hayles, N.K. (2006). 'Traumas of Code'. *Critical Inquiry*. **33**: 136-157.

Hockey, S.M. (2000a). 'Dictionaries and Lexical Databases'. *Electronic texts in the humanities: Principles and Practice*. Oxford; New York: Oxford University Press, pp. 146-171.

Hockey, S.M. (2000b). *Electronic texts in the humanities: Principles and Practice*. Oxford; New York: Oxford University Press.

Knowles, F.E. (1989). 'Computers and Dictionaries'. *Wörterbücher: ein*

internationales Handbuch zur Lexikographie. Hausmann, F.J., Reichmann, O., Wiegand, H.E. (eds.). Berlin; New York: W. de Gruyter, pp. 1645-1672.

Ryszard Kruk S., McDaniel B. (eds.) (2009). *Semantic Digital Libraries.* Berlin: Springer.

Lemberg, I., Schröder, B., Storrer, A. (eds.) (2001). *Chancen und Perspektiven computergestützter Lexikographie: Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher.* Tübingen: M. Niemeyer.

Maness, J.M. (2006). 'Library 2.0 Theory: Web 2.0 and Its Implications for Libraries'. *Webology.* **2**.

Meijs, W. (1992). 'Computers and Dictionaries'. *Computers and Written Texts.* Butler, C. (ed.). Oxford; Cambridge, Mass.: Blackwell, pp. 141-165.

Miller, G.A. et al.. 'Introduction to WordNet: An Online Lexical Database'. *Five papers on WordNet.*

Miller, P. (2005). 'Web 2.0: Building the New Library'. *Ariadne.* **45**.

Nielsen, S. (2009). 'Reviewing printed and electronic dictionaries: A theoretical and practical framework'. *Lexicography in the 21st Century. In honour of Henning Bergenholtz.* Nielsen, S., Tarp, S. (eds.). Amsterdam: John Benjamins, pp. 23-41.

Novotny, E. (2006). *Assessing Reference and User Services In a Digital Age.* Binghamton, NY: Haworth Information Press.

Rundell, M. (2009). 'The future has arrived: a new era in electronic dictionaries'. *MED Magazine.* **54**.

Sinclair, J. (2000). 'Lexical Grammar'. *Darbai ir Dienos.* **24**.

Stamou, S., et al. (2002). 'BALKANET A Multilingual Semantic Network for the Balkan Languages'. *Proceedings of the International Wordnet Conference.* Mysore, India, 21-25 January 2002, pp. 21-25.

Tasovac, T. (2008). 'Why not every picture is worth a thousand words: digital libraries from a textual perspective'. *Proceedings of the International Conference "Electronic Libraries - Belgrade 2008".* University of Belgrade, 25-27 September 2008.

Tasovac, T. (2009). 'More or Less than a Dictionary: WordNet as a Model for Serbian L2 Lexicography'. *Infoteka - Journal of Informatics and Librarianship.* **10**: 13a - 22a.

Tufis, D., Cristea, D., Stamou, S. (2004). 'BalkaNet: Aims, Methods, Results and Perspectives. A General Overview'. *Science and Technology.* **7**: 9-43.

Vossen, P. (ed.) (1998). *EuroWordNet: A Multilingual Database with Lexical Semantic Networks.* Dordrecht, The Netherlands; Boston, Mass.: Kluwer Academic.

Шведова, Н. (1988). 'Парадоксы словарной сатьи'. *Национальная специфика языка и ее отражение в нормативном словаре. Сборник статьей.* Караулов, Ю.Н. (ed.). Москва: Наука, pp. 6-11.

Notes

1. The first publicly available dictionary application programming interface was made available by the Wordnik project in October 2009. See `http://api.wordnik.com/signup/`.