The politics of code. How digital representations and languages shape culture

Domenico Fiormonte [1],*, Desmond Schmidt [2], Paolo Monella [3], Paolo Sordi [4]**

[1] Dipartimento di Scienze Politiche / Via Chiabrera 199 – 00145 Rome, Italy

[2] eResearch ITEE University of Queensland, Australia

[3] University of Palermo, Department of "Culture e Società", Viale delle Scienze, Edificio 12, Palermo, Italy

[4] University of Rome Tor Vergata, Department of Humanistic Studies, Via Columbia 1, Rome, Italy

E-Mails: domenico.fiormonte@uniroma3.it (D.F.); desmond.allan.schmidt@gmail.com (D.S.); paolo.monella@unipa.it (P.M.); paolo.sordi@uniroma2.it (P.S.)

* Author to whom correspondence should be addressed; E-Mail: domenico.fiormonte@uniroma3.it; Tel.: +39-06-5733-5372; Fax: +39-06-5733-5280.

Introduction

Digital languages and instruments are not only powerful tools for simplifying and enhancing the work of humanists and social scientists, they also create new cultural representations and self-representations that transform both the epistemology and the practice of research.

In particular, digital representations can influence and shape our cultural artefacts and everyday experiences in various ways. For example, we are influenced in the way we represent everyday digital objects, such as the Windows folder as a metaphor for 'document box'. This is at the level of the interfaces that 'produce users through benign interactions […]. That is, as ideology creates subjects, interactive and seemingly real-time interfaces create users who believe they are the "source" of the computer's action' ([1], pp. 66-68). We are also influenced in the ways we write software or encode a document through specific languages (e.g. Python or HTML). However, it is easy to demonstrate that such a distinction is artificial (and often damaging). From a semiotic point of view, both representations (visible interface and invisible coding) are 'modelling systems' (2) that cast their influence on overlapping political, social, cognitive and epistemological domains. In the case of the interface we are

talking about an influence mainly on practices and processes (social and cognitive dimensions), and in the coding case we are dealing with the theory and interpretation of information structures (linguistic, hermeneutical and epistemological dimensions).

In our paper we will focus on this second aspect, and show some examples of how code and encodings are shaping the way we conceive and practise the work of reconstruction, conservation and representation of information structures and cultural artefacts. To this aim, we will discuss three encoding tools widely used in the Humanities and Social Sciences communities: HTML, the de facto standard for encoding Web documents and pages, Unicode, an industry standard designed to allow text and symbols from all of the writing systems of the world to be consistently represented and manipulated by computers, and XML (eXtensible Markup Language), which defines a set of rules for encoding documents.

HTML War

The evolution of HTML (Hypertext Markup Language) as a structural language and application standard of the World Wide Web has been largely shaped by forces outside of the World Wide Web Consortium (W3C), the international community assembled by Tim Berners-Lee 'to lead the World Wide Web to its full potential by developing protocols and guidelines that ensure the long-term growth of the Web' (http://www.w3.org/Consortium/).

Even before it became a W3C Recommendation on October 28, 2014, HTML5 had already reached in 2009 the status of a de facto standard on the Web. This was promoted and encouraged by the Web Hypertext Application Technology Working Group (WHATWG), a community established by Apple, the Mozilla Foundation and Opera Software (joined later by Google) that was in open opposition to the development of the language of the Web as envisioned by the W3C and 'concerned about the W3C's direction with XHTML, lack of interest in HTML and apparent disregard for the needs of real-world authors.' (https://wiki.whatwg.org/wiki/FAQ#What_is_the_WHATWG.3F)

Just as in the days of the browser wars between Netscape Navigator and Microsoft Internet Explorer (4), tensions collide and versions of code for the Web are released that appear to be inspired by needs of the Web authors and users, but in reality are tied to strategies for controlling a market enhanced by Web 2.0 apps, social media and the eternal promise of the Semantic Web (5).

Controlling the development of HTML means controlling the competition between the different applications used to access the Web (5); managing the scheduling and release of mobile 'apps' that can now compete at the same level as their desktop rivals, without being forced to adopt or implement competing solutions; and to determine how texts and information on the Web can be searched and interrelated, making the definitive migration from a Web based on document search to one based on the collection of user data. In other words, behind the HTML war there is a growing obsession by industry and governments with our digital traces, habits and personal information (6) (7).

The cultural and political biases of XML

A markup language provides a means to formally describe a text's structure, and to analyse its data in depth. Its usefulness will be in proportion to how much information it can set out, include and preserve. The word 'markup' itself shows its original bias: ironically, like other computing languages, XML is one the most faithful successors to the Gutenberg model; its basic aim is to imitate and preserve

structured information as laid out in modern printed books. But is hierarchical and structured information an inherent and universal feature of texts (let alone writing)? The bias inherited from print has forced us to think of a text as a stable product, but if we either look at the different historical representations of a given text or at its documented writing stages, it is clear that there is not one text, but as many different texts as there are mechanisms of writing, material production, intertextual paths and methodologies of reconstruction (8). Not only there is a potential conflict between the linear and hierarchical nature of current markup languages, and the intrinsic dynamic nature of the writing process, current text encoding tools and methodologies seems to constitute the most serious obstacle to the development of an independent theory of digital text.

A typical example of the overlapping between geopolitical settings and technical choices can be found in organisations like the TEI (Text Encoding Initiative [9]), an international consortium that defines guidelines for encoding cultural heritage documents in XML. However, the practice of defining encoding standards to allow electronic documents to be processed ultimately by shared software at the presentation level, does not work nearly as well for the encoding of historical primary sources (10). Attempts to declare 'standard' names for textual features overlook significant variations in the interpretation, selection and application of those codes by different groups, individuals and cultures. Once again, the 'practice' of code is the result of the 'theory' reflected by certain groups and interests. Encoding tools and methodologies become examples of 'symbolic capital' (11) used by the TEI community to export a universalising, western-centric approach to the representation of cultural artefacts.

Universalizing the typography: Unicode

The Unicode standard aims at constituting a universal and inclusive mapping of all graphemes from existing and past writing systems. Yet, some assumptions that underlie the standard are shaped by the culture-specific standardisation of the graphical representation of language generated by Gutenberg's invention of moveable type printing.

The Unicode writing system, now forced upon all scripts, is based on a discrete ordered sequence of individual characters flowing only in one direction and dimension. However, in many writing systems based on handwriting, including contemporary Arabic and some Indic scripts, some graphemes 'orbit' around a central grapheme and can be written not only after (left or right, depending on the main direction of the script), but also before, above or under the 'main' grapheme. This is the case of ancient Greek 'hypogegrammenon (subscribed) iota', written before another vowel if lowercase, but after that vowel if uppercase. A wider case study is provided by the right-to-left Indic scripts where some vowels, although pronounced after a consonant, may be written before it ([12], chapter 12.1; [13]; [14], par. 6.3).

Another assumption underlying the Unicode model is a one-to-one correspondence between phoneme and grapheme. The European medieval handwriting conventions based on the Greek or Latin alphabets are an interesting case study in this respect: ligatures, brevigraphs (sometimes represented by apparent diacritics such as macrons) and logographs are systematic and 'regular' here, and the underlying model is incompatible with the Unicode/Gutenberg one-to-one correspondence between a phoneme and a character in a linear sequence. However, the XML/TEI Guidelines for the transcription of pre-modern primary sources simply recommend the use of Unicode for this purpose (TEI P5 Guidelines, chapter 5).

In all these cases, the current version of Unicode imposes the Gutenberg model (based on the Latin alphabet used in Europe in the Modern Age) upon writing systems based on different models, and all

discrepancies are treated as 'exceptions' to be dealt with at the presentational rendering level through software workarounds.

Conclusions

These case studies show that digital 'standards' always reflect a cultural bias (15), and that the level of encoding is never neutral, but tends to assume (and overlap with) universalising discourses that are usually invisible at the surface of technology (16).

Digital Universalism (17) is deeply intertwined with the question of language, which in its turn controls our encoding practices. Code imperialism and linguistic imperialism (18) are two sides of the same coin. We are facing a codex universalis, based on the English lingua franca, used to articulate and exert a power of standardisation and control.

But codes and encoding(s) before or along 'semiotic', 'hermeneutic' or 'cultural' are also political phenomena. Looking at boards, offices and steering committees reveals that consortia, associations and organisations like Unicode, ICANN, TEI or the W3C are informed by Anglophone hegemonies, and their decisions are shaped by their commercial interests, ideologies and cultures. So apparently 'neutral', 'technical' decisions, as can be observed in Unicode, TEI or other organisations, tend to oversimplify and standardise the complex diversity of languages and cultural artefacts.

As Friedrich Kittler put it: 'Codes—by name and by matter—are what determine us today, and what we must articulate if only to avoid disappearing under them completely. … Today, technology puts code into the practice of realities, that is to say: it encodes the world' ([19], p. 40). So, is it still in our power to code (encode reality), or rather is code imposing on us its biases and constraints?

In the conclusion of our presentation we will try to propose solutions to avoid or reduce the impact of universalistic encoding, and report on alternative experiences and experiments that try to resist the effects of 'colonial computing' (20).

References

1. Chun, H.K.W., *Programmed visions. Software and memory*, MIT Press: Cambridge (MA).
2. Uspenskij, B. A.; Ivanov, V. V.; Piatigorskij, A.; Lotman, J. M., Tezisy k semiotičeskomu izučeniju kul'tur (v primenenii k slavjanskim tektstam). In *Semiotyka i Struktura Tekstu*. Studia święcone VII międz. Kongresowi Slawistów; Mayenowa, M. R. (Ed.), Warszawa, pp. 9-32. Trad. en. "Theses on the Semiotic Studies of Culture (As applied to Slavic Texts)". In van der Eng, J., Grygar, M., Eds.; *Structure of texts and semiotics of culture*, Mouton: The Hague, 1973; pp. 1-28.
3. Sordi, P. Leggere il codice del Web. *Testo e Senso*, **2010**, 11. Available online: http://testoesenso.it/article/view/1 (accessed on 6 May 2015).
4. Zeldman, J., Marcotte, E. *Designing with Web Standards*, 3rd edition; New Riders: Berkeley, USA, 2010.
5. Ford, P. On HTML5 and the Group That Rules the Web. *The New Yorker*, November 20, 2014. Available online: http://www.newyorker.com/tech/elements/group-rules-web (accessed on 6 May 2015).
6. Bauman, Z., Lyon, D. *Liquid Surveillance. A Conversation*. Polity Press: Cambridge, UK, 2013.
7. Bowker, G. Y. Data Flakes: an afterword to 'Raw Data' is an oxymoron. In Gitelman, L. (Ed.) *'Raw Data' is an oxymoron*. MIT Press: Cambridge (MA), USA, 2013, pp. 167-171.

8. Fiormonte, D., Martiradonna, V., Schmidt, D. Digital Encoding as a Hermeneutic and Semiotic Act: The Case of Valerio Magrelli. *Digital Humanities Quarterly* **2010**, 4, 1. Available online: http://digitalhumanities.org/dhq/vol/4/1/000082/000082.html (accessed on 7 May 2015).

9. TEI Consortium, Eds. T*EI P5: Guidelines for Electronic Text Encoding and Interchange*. TEI Consortium, 2007. Available online: http://www.tei-c.org/Guidelines/P5/ (accessed 6 May 2015).

10. Schmidt, Desmond. Towards an Interoperable Digital Scholarly Edition. *Journal of the Text Encoding Initiative* **2014**, 7. Available online: http://jtei.revues.org/979 (accessed on May 6 2015).

11. Bourdieu, P. *Distinction: A social critique of the judgement of taste*. Harvard University Press: Cambridge (MA), USA, 1984.

13. Perri, A. Al di là della tecnologia, la scrittura. Il caso Unicode. *Annali dell'Università degli Studi Suor Orsola Benincasa*, **2009**, II, 725-748.

14. Constable, P. G. Understanding Unicode: a general introduction to the Unicode Standard. *NRSI: Computers & Writing*, 2001. Available online: http://scripts.sil.org/cms/scripts/page.php?item_id=IWS-Chapter04b (accessed 7 May 2015).

15. Carey, J. W. *Communication as Culture. Essays on Media and Society*. Revised Edition. Routledge Routledge: New York and Abingdon, 2009.

16. Galloway, A.R. *The interface effect*. Polity Press: London, 2012.

17. Chan, A. S. *Networking Peripheries: Technological Futures and the Myth of Digital Universalism*. MIT Press: Cambridge (MA), 2014.

18. Phillipson, R. *Linguistic imperialism continued*. Routledge: New York and London, 2009.

19. Kittler, F. Code (or, How You Can Write Something Differently). In *Software Studies: A Lexicon*. Fuller, M., Eds.; MIT Press: Cambridge (MA), 2008, pp. 40-47.

20. Ali, M. Towards a decolonial computing. In *CEPE 2013: Computer Ethics: Philosophical Enquiry*, 1-3 July 2013, Lisbon, Portugal, International Society of Ethics and Information Technology, 2014, pp. 28–35.