Croatiae auctores Latini

Neven Jovanović

August 2, 2012

The neo-Latin writings of people from Croatia are today arguably seen as part of national cultural heritage.[1] The much abused word "arguably" is appropriate here because, for my statement to be true, it is not enough just to "have" or "preserve" these writings. We have also to claim them, to learn to know them, in short — we have to do something with them. One way of "doing something" is the digital collection *Croatiae auctores Latini*, of which I am one of the editors, and which I'll today present both as an artefact and as a research challenge.

Allow me, first, to give some background information. In lands which form today's Croatia, for a number of cultural and political reasons Latin served as an administrative and literary language during more than a thousand years, from 850 until 1900. A bibliographical project has shown that between 1474 and 1848 there were printed 6123 Latin titles (by two-thirds more than there were titles printed in Croatian) containing texts by 3882 authors somehow connected with Croatia. We may therefore estimate that the thousand years of Croatian Latin produced at least 7000 works by 4000 authors. Forty years ago, in 1969-1970, a tiny fraction of these authors and works — 37 authors, 116 texts — was selected for presentation in the influential anthology by Veljko Gortan and Vladimir Vratović, *Hrvatski latinisti – Croatici auctores qui Latine scripserunt*. Judging by recent reference works, the actual "canon" of Croatian Latin includes somewhere between 100 and 200 authors and some 70 individual titles. That is one percent of estimated total works, 3 to 5 percent of authors.

1

The digital collection *Croatiae auctores Latini* (CroALa) was first published in 2009.[2] The collection has an ISBN and an international editorial board. It is fully and freely accessible and searchable over the internet, and all texts can be downloaded and reused (under a so-called Creative Commons

---

[1]Illustrative material for this paper is available at http://www.ffzg.unizg.hr/klafil/dokuwiki/doku.php/z:croala-ianls.

[2]See http://www.ffzg.unizg.hr/klafil/croala.

license) from the CroALa SourceForge site. The texts are encoded semantically rather than presentationally – for example, a `l` tag means "this is a line of poetry", regardless of how it is laid out on page or on screen. For encoding, we use the Text Encoding Initiative TEI XML standard. This format, accepted widely in the humanities, makes it possible to mark up practically every textual feature we encounter, and ensures exchange and independence of data. The texts are published (that is, made searchable over the internet) through PhiloLogic, a full-text search, retrieval and analysis tool developed at the University of Chicago.

At the moment the CroALa collection contains 236 documents (a wider category of "document", such as a book of poetry, can comprise several texts) by 131 authors, and about 3,340,000 words. The authors represented are the well known ones, such as those mentioned in IJsewijn's *Companion*, as well as others, no less important or intriguing – such as Pavao Pavlović, the chronicler of 14th century Zadar; Antun Rozanović, who wrote a history of the Turkish attack on the island of Korčula in 1571; Kajetan Vičić, a 17th century author of the longest Croatian neo-Latin epic poem, the *Jesseid* (13,523 verses, 84,448 words); Balthasar Adam Krčelić, who left a voluminous, posthumously censored *chronique scandaleuse* of the 18th century Croatia, the *Annuae*; and curious Latin inscriptions and postcards of don Frane Bulić, archaeologist of Salona near Split.

2

The advantages of having all these texts in one place, and made searchable not only by volume and page, but also by individual words or expressions — and searchable conveniently from our study-room — all this is, I guess, self-evident. Imagine a library where books are not separated by book-covers, where we can read not only linearly, but also, as Woody Allen would say, diagonally – and where any passage we're interested in can be retrieved extremely quickly. However, there arises another question, which is a central question of the so-called digital humanities: what can be done digitally, computationally, that is not only "more of the same that we usually do, only quicker"? What can we do with CroALa that would be impossible without it?

Before I offer an answer, it is important to point out what CroALa is not, or isn't yet. It is not a linguistic corpus, a carefully balanced and representative language sample. It is, obviously, also not a collection of *all* Croatian neo-Latin. It is also not an anthology of "the best" in Croatian neo-Latin. At the moment it is uneven by value, period, genre, and size. It has necessarily been grounded on a set of interpretations about what is, or isn't, Croatian neo-Latin, and what is, or isn't, epic poetry or elegy or epigram (we tried to keep these interpretations as modest and as common as possible; I hope our Hungarian and Italian colleagues won't have many objections).

What is CroALa, then? It is a laboratory. It is an instrument that en-

ables us to form hypotheses about neo-Latin writings — and to test them, and refine them into other hypotheses. Moreover, this can be done quickly and easily; thanks to PhiloLogic, finding all superlatives among three million words takes less than a second. This also means that our experiments can be repeated by other people, that the validity of our findings can be tested. Even more importantly, this means that we can experiment with ideas that would otherwise simply be too expensive — too time-consuming, and too tedious — to follow. And, finally, this means that little by little we start having ideas which are, in their Procrustean bed of computational explicitness and precision, slightly different from the ones we usually have.

Let me then, in the rest of the time for my paper, illustrate these different ideas with four examples. These are not the final results nor anything you would be ready to publish in a journal; I offer here four starting points for investigation and inspiration.

3

3.1

The first task is simple: we want to find in CroALa all passive infinitives ending in *-ier* and look at their distribution. Simplicity is deceptive, because we have to get rid of words such as *mulier*, and the occasional foreign name or title (such as *vizier*). However, once we clean up our list, PhiloLogic finds 541 occurrences of 245 infinitives immediately. The interesting fact is that all these infinitives occur in *poetry*; the single exception, in a 1506 letter by Ilija Crijević, actually contains verse written as prose.

3.2

With the next experiment we wanted to demonstrate the large-scale querying possibilities of CroALa. Following the suggestion of Hans Helander and others that the words which were rare in classical Latin usage were used without hesitation, even often, in neo-Latin, we compiled a list of words considered rare in OLD and Lewis – Short. The list is far from perfect; however, it contains impressive 7974 entries (think about it: if a search took just one second, we would need more than two hours just to go through the list). Using a simple programming script, actually a list of commands, we sent these 7974 words (slightly transformed for spelling variants) to CroALa, and noted down the results. You will notice that the five- and six-letter queries produce unimpressive results, and should be improved, but occurrences of the longer strings are well worth studying.

3.3

The opposite of large-scale searches happens when we look for something specific. For example, when we want to see which texts in CroALa mention a mythological figure, or figures – in our case, the Danaides. Obviously, similar searches can be done on a name (Dubrovnik, the Turks), or on a theme (such as *somnus*, *amor*). The search has formed a nice cluster of poets — they all use the Danaides in their descriptions of the underworld, and the descriptions and the poems await our close readings.

3.4

Finally, I took a finding from another study — in this case, from Hans Helander's impressive book on 17th century Swedish neo-Latin — and tested it on our collection: is the motif of the *theatrum mundi* as important for Croatian neo-Latin texts in CroALa as it is for Helander's corpus? First results point in the other direction.

4

To conclude. Digital neo-Latin research faces currently a double challenge. We have to make use of knowledge, methods, and ideas being developed by the so-called digital humanities — and we have to use it in a way that our more traditionally minded colleagues would find interesting. It is not enough to produce tools and collections; we have to *question* them at the same time. What can be done with it, what cannot? For example, it is easy to research figures of speech in CroALa — but figures of thought, such as antithesis or hyperbole, can be accessed only indirectly and imperfectly. On the other hand, everything we find in the collection, everything we conclude from the findings, can be tested and verified — while we often have to take the insights of Helander at his word (synecdoche is the fate of every printed book), the search boxes and links of CroALa and similar systems invite everybody to experiment. So, think about your own search terms; test your theses on our texts; bring your own texts — even just one or two — and turn them into a PhiloLogic database. And then let's find out what can be done.