Unveiling private anthologies: Exploring lay readers' poetic repertoire
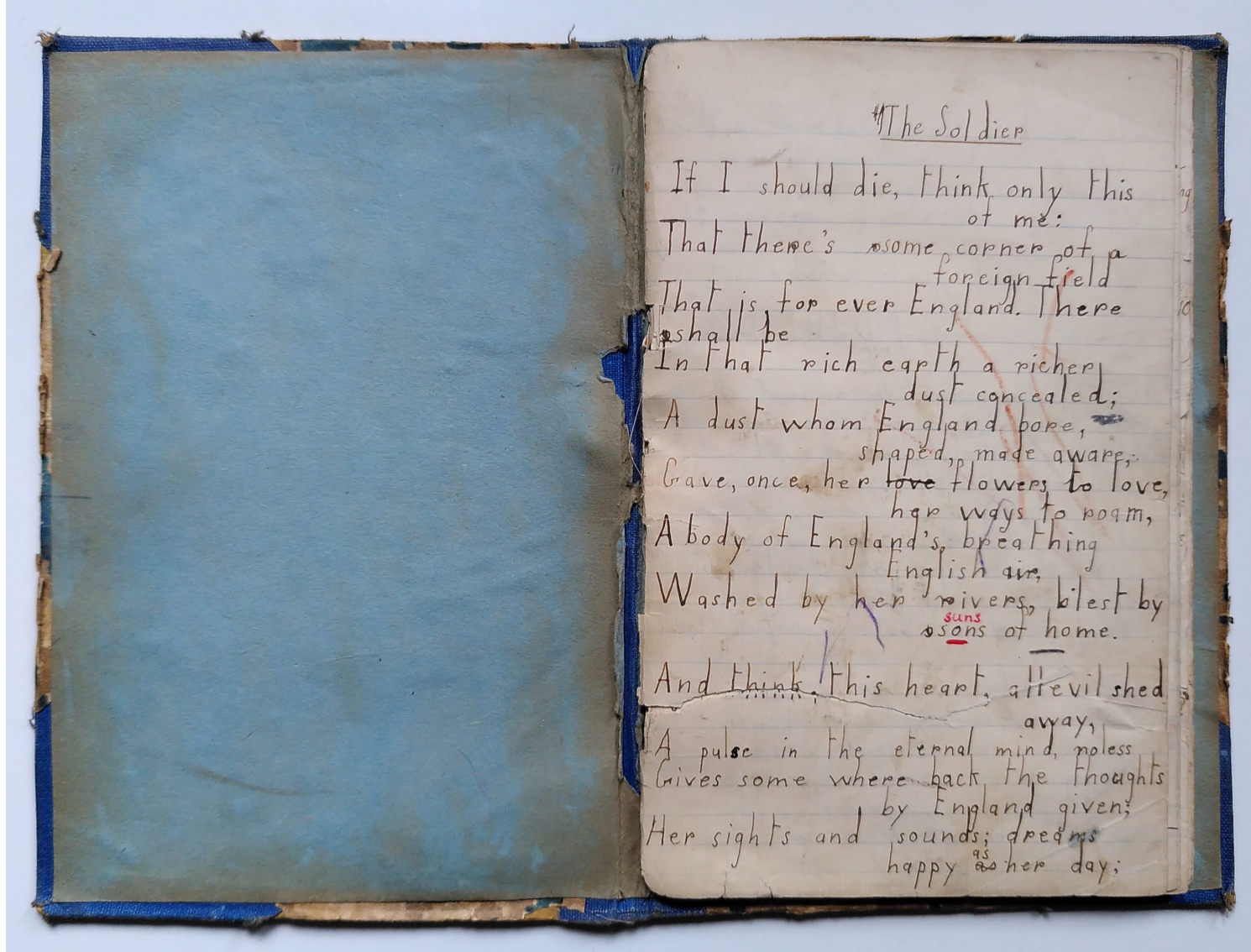
Álvaro CEBALLOS VIRO (Université de Liège)

PRIVATE ANTHOLOGIES

For well over two centuries, lay readers have been jotting down in notebooks poems they considered emotionally or epistemologically relevant.

Characteristics :
- different periods and authors;
- usually plurinational, often plurilingual;
- combine canonical and popular, sometimes anonymous poems;
- the moment of transcription often remains undetermined.

They are a fascinating object for studying the empirical uses of poetry! Could it be that, beyond sharing generic, material, and formal characteristics, they also share semantic features?

Hypothesis: private anthologies exhibit specific semantic networks, offering insights into lay readers' poetic repertoires.

CORPUS

Experimental corpus:
5 notebooks containing private poetic anthologies written in Spain between *ca.* 1850 and *ca.* 1940.
Transcribed with Transkribus' 'Coloso Español' (4,8% CER, subsequently revised), normalized and lemmatized with Python (Spacy, es_core_news_sm).

Control corpus:
22 poetic works published in Spain during the same century, including five commercial lyric poetry anthologies.

A most frequent words analysis with Voyant show no difference between experimental and control group.

		Términos		Contar	Tendencia
⊞ ☐	1	amor	love	2218	
⊞ ☐	2	alma	soul	1837	
⊞ ☐	3	vida	life	1551	
⊞ ☐	4	luz	light	1308	
⊞ ☐	5	ojos	eyes	1304	
⊞ ☐	6	dios	god	1258	
⊞ ☐	7	cielo	heaven / sky	1234	
⊞ ☐	8	corazón	heart	1119	
⊞ ☐	9	oh	oh	1115	
⊞ ☐	10	sol	sun	1089	
⊞ ☐	11	mundo	world	1047	
⊞ ☐	12	son	are	1042	
⊞ ☐	13	tierra	earth	1031	
⊞ ☐	14	siempre	always	995	
⊞ ☐	15	día	day	969	
⊞ ☐	16	ella	she	932	

TOPIC MODELING

Topic modeling techniques could reveal thematic differences. In topic modeling, word co-occurrence can be interpreted as a proxy for semantic affinity, similar (but not identical) to what is commonly referred to as a 'theme'.
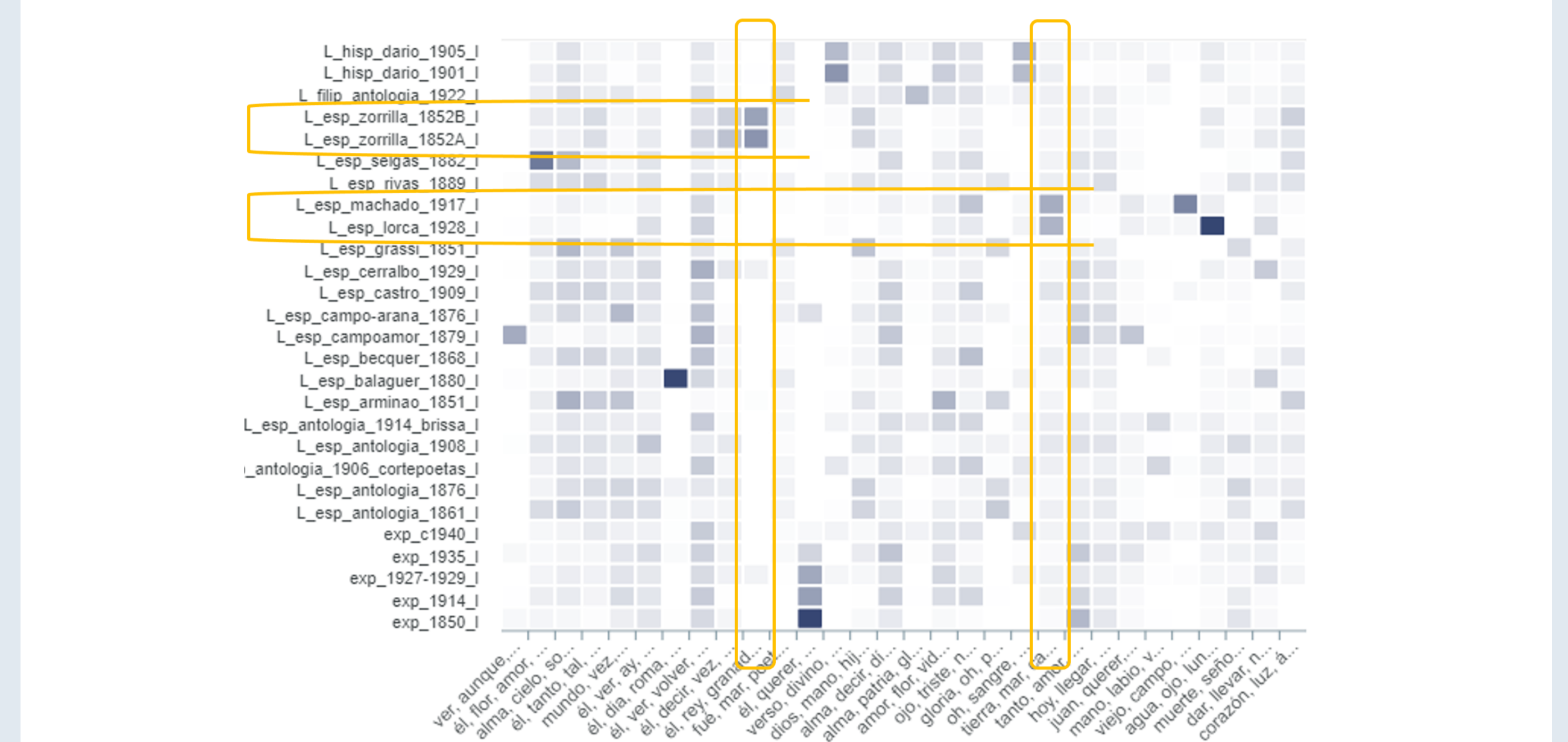
Topic modeling reveals recurring topics in the control corpus (see heatmap 1):

- José Zorrillas (†1893) arabizing poems: 'king,' 'moor,' 'tower,' 'Granada,' 'faith,' 'Christian,' etc.
- Antonio Machado (†1939) and Federico García Lorca (†1936) share the 'elemental hiker' topic ('earth,' 'moon,' 'path,' 'field,' 'night,' 'air,' 'wind,' 'water,' 'fountain,' 'river,' 'sea'…).
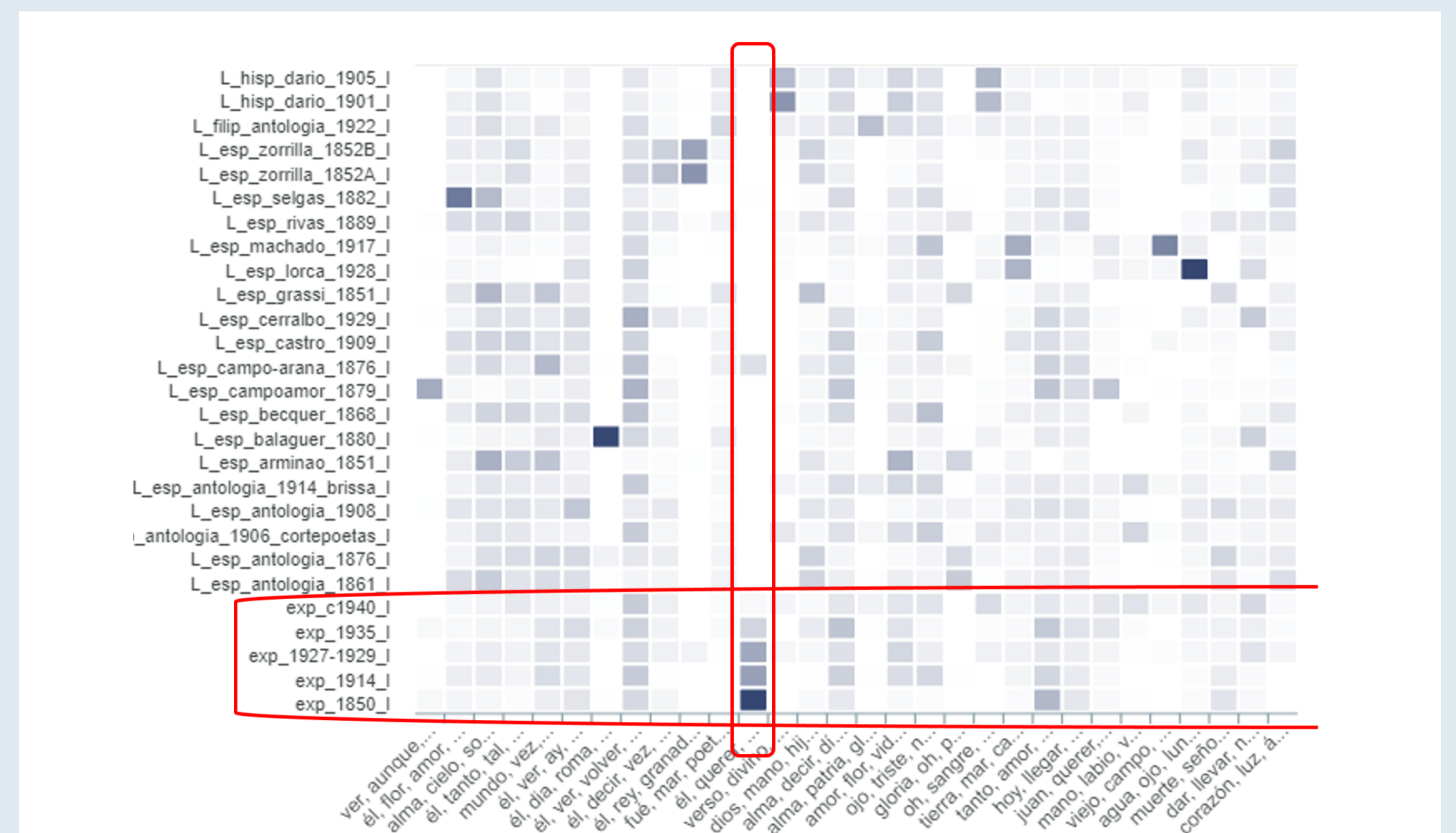
I tried out DARIAH-Topics Explorer (LDA) with changing iterations and numbers of topics. By increasing the number of topics to a reasonable threshold, these trends become more pronounced.

Both heatmap 1 (above) and heatmap 2 (below) display 30 topics after 300 iterations.

20t, 100i	20t, 300i	20t, 1000i	30t, 300i	30t, 1000i
love	love	love	love	love
love/want	love/want	love/want	love/want	love/want
heart	heart	heart	heart	so
see	woman	woman	woman	woman
lose	life	life	lover	life
god	god	god	(3rd p.s.)	(3rd p.s.)
owe/duty	flower	flower	see	see
give	always	always	give	day
soul	live	live	watch	soul
chest/breast	chest/breast	chest/breast	chest/breast	watch
say	say	say	say	happen
memory	a lot	a lot	a lot	a lot
let	think	think	sad	think
worship	sweet	sweet	sweet	until
although	die	die	evil/bad	die

Words composing the specific topic of private anthologies, with varying topic numbers and iterations.

As heatmap 2 shows, **private poetic anthologies also present an exclusive topic and it contains many of the most frequent words in the total corpus:** 'love,' 'heart,' 'god,' 'life/live,' 'soul,' 'woman/she,' 'eyes/see'.

In the control corpus, the most frequent words do not establish links among themselves. Conversely, in private poetic anthologies, those words *do* interact, attracting each other to form a topic where love and spirituality intertwine, akin to the medieval songbooks and the *religio amoris*, a hyperbolic codification of romantic relationships (notice also 'worship,' 'die,' 'a lot'...). This would be consistent with several signposts that situate private anthologies in contexts of seduction.

FUTURE DEVELOPMENTS

Further research would require to:
- objectify and better articulate the likelihood function in this kind of analysis;
- consider Dirichlet parameters in assessing the importance of each topic;
- analyze the variations between original and handwritten poems;
- expand the corpus of lay readers' anthologies, regardless of their language of writing;
- try out a taxonomy of social and individual uses, paying attention to gendered practices.

SELECTED REFERENCES

- CERTEAU, Michel de (1990 [1980]): *L'Invention du quotidien. I, Arts de faire*, Paris: Gallimard.
- EDDY, Matthew D. (2018): «The Nature of Notebooks: How Enlightenment Schoolchildren Transformed the Tabula Rasa», in *Journal of British Studies*, 57, 275-307.
- ESCOLANO, Agustín (2000): «La copie de manuscrits comme pratique scolaire d'appropriation et ses usages de l'écriture», in: Ch. Barré-De Miniac (ed.): *Copie et modèle: usages, transmission, appropriation de l'écrit*, Nancy: Institut National de Recherche Pédagogique, 41-56.
- HENNY, Ulrike *et al.* (2016): «Topic, Genre, Text. Topics im Textverlauf von Untergattungen des spanischen und hispanoamerikanischen Romans (1880-1930)», DHd-Tagung, Univ. Leipzig: http://www.dhd2016.de/abstracts/vortr%C3%A4ge-055.html
- HEESEN, Anke te (2005): «The Notebook: a Paper Technology», in B. Latour / P. Weibel (eds.): *Making Things Public. Atmospheres of Democracy*, Cambridge / London: The MIT Press, 582-589.
- JELODAR, Hamed *et al.* (2019): «Latent Dirichlet Allocation (LDA) and topic modeling: models, applications, a survey», in *Multimedia Tools and Applications*, 78, 15169-15211
- MINZETANU, André (2016): *Carnets de lecture. Généalogie d'une pratique littéraire*, Vincennes: Presses Universitaires de Vincennes.
- VAYANSKY, Ike / KUMAR, Sathish A. P. (2020): «A review of topic modeling methods», *Information Systems*, 94.
- WALTER, Stefan (2019): *Der Staat und die Werte: Wertwandel in Poesiealben in der DDR und Bundesrepublik 1949–1989*, Wiesbaden: Springer VS.

ACKNOWLEDGEMENTS